PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 :		(11) International Publication Number:	WO 89/ 01944
C07H 15/12, C12Q 1/70, 1/02 C12Q 1/06, C12P 21/00, 19/34 C12P 1/04, C12N 15/00, 7/00 C07K 13/00, 3/00, A61K 37/68 A61K 39/00, 45/02	A1	(43) International Publication Date:	9 March 1989 (09.03.89)

(21) International Application Number: PCT/US88/02940

(22) International Filing Date: 1 September 1988 (01.09.88)

(31) Priority Application Numbers: 094.322 141.649

(32) Priority Dates: 4 September 1987 (04.09.87) 7 January 1988 (07.01.88)

(33) Priority Country: US

(60) Parent Applications or Grants
(63) Related by Continuation
US 094.322 (CIP)
Filed on 4 September 1987 (04.09.87)
US 141.649 (CIP)
Filed on 7 January 1988 (07.01.88)

(71) Applicant (for all designated States except US): BIOGEN, INC. [US/US]: 14 Cambridge Center, Cambridge, MA 02142 (US).

(72) Inventors: and
(75) Inventors/Applicants (for US only): FISHER, Richard, A. [US/US]: 1079 Beacon Street, Brookline, MA 02146 (US). GILBERT, Walter [US/US]: 107 Upland Road, Cambridge, MA 02140 (US). SATO, Vicki, L [US/US]: 43 Larch Road, Cambridge, MA 02138 (US). FLAVELL, Richard, A. [GB/US] 182 Reservoir Road, Killingworth, CT 06417 (US). MARAGANORE, John, M. [US/US]: 84 Patrick Road, Tewksbury, MA 01876 (US). LIU, Theresa, R. [US/US]: 102 Emerson Road, Milton, MA 02186 (US).

(74) Agents: HALEY, James, F., Jr. et al.; Fish & Neave, 875 Third Avenue, New York, NY 10022-6250 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), BJ (OAPI patent), CF (OAPI patent), CH (European patent), CG (OAPI patent), CM (OAPI patent), DE (European patent), DK, FR (European patent), GA (OAPI patent), GB (European patent), IT (European patent), JP, KR, LU (European patent), ML (OAPI patent), MR (OAPI patent), NL (European patent), NO, SE (European patent), SN (OAPI patent), TD (OAPI patent) TG (OAPI patent), US.

(54) Title: DNA SEQUENCES, RECOMBINANT DNA MOLECULES AND PROCESSES FOR PRODUCING SOLUBLE T4 PROTEINS

(57) Abstract

This invention relates to DNA sequences, recombinant DNA molecules and processes for producing soluble T4 protein. More particularly, this invention relates to DNA sequences that are characterized in that they code on expression in an appropriate unicellular host for soluble forms of T4, the receptor on the surface of T4+ lymphocytes, or derivative thereof. In accordance with this invention, the DNA sequences, recombinant DNA molecules and processes of this invention may be employed to produce soluble T4 essentially free of other proteins of human origin. This soluble protein may then advantageously be used in the immunotherapeutic and diagnostic compositions and methods of this invention. The soluble T4-based immunotherapeutic compositions and methods of this invention are useful in treating immunodeficient patients suffering from diseases caused by infective agents whose primary targets are T4+ lymphocytes. According to preferred embodiment, this invention relates to soluble T4-based compositions and methods which are useful in preventing, treating or detecting acquired immune deficiency syndrome, AIDS related complex and HIV infection.

1

DNA SEQUENCES, RECOMBINANT DNA MOLECULES AND PROCESSES FOR PRODUCING SOLUBLE T4 PROTEINS

TECHNICAL FIELD OF INVENTION

This invention relates to DNA sequences, recombinant DNA molecules and processes for producing soluble T4 proteins. More particularly, this invention relates to DNA sequences that are characterized in that they code on expression in an appropriate unicellular host for soluble forms of T4, the receptor on the surface of T4$^+$ lymphocytes, or derivatives thereof. In accordance with this invention, the DNA sequences, recombinant DNA molecules and processes of this invention may be employed to produce soluble T4 essentially free of other proteins of human origin. This soluble protein may then advantageously be used in the immunotherapeutic, prophylactic, and diagnostic compositions and methods of this invention.

The soluble T4 protein-based immunotherapeutic compositions and methods of this invention are useful in treating immunodeficient patients suffering from diseases caused by infective agents whose primary targets are T4$^+$ lymphocytes. According to a preferred embodiment, this invention relates to soluble T4 protein-based compositions and methods which ar useful in preventing, treating or detecting

acquired immune deficiency syndrome, AIDS related complex and HIV infection.

BACKGROUND ART

5   The class of immune regulatory cells known as T cell lymphocytes can be divided into two broad functional classes, the first class comprising T helper or inducer cells -- which mediate T cell pro- liferation, lymphokine release and helper cell inter- actions for Ig release, and the second class compris-
10   ing T cytotoxic or suppressor cells -- which parti- cipate in T cell-mediated killing and immune response suppression. In general, these two classes of lymphocytes are distinguished by expression of one of two surface glycoproteins: T4 (m.w. 55,000-62,000
15   daltons) which is expressed on T helper or inducer cells, probably as a monomeric protein, or T8 (m.w. 32,000 daltons) which is expressed on T cytotoxic or suppressor cells as a dimeric protein.

       The primary structures of T4 and T8 have
20   been deduced from their respective cDNA sequences [P. J. Maddon et al., "The Isolation and Nucleotide Sequence Of A cDNA Encoding The T Cell Surface Protein T4: A New Member Of The Immunoglobulin Gene Family", Cell, 42, pp. 93-104 (1985); D. R. Littman et al.,
25   "The Isolation And Sequence Of The Gene Encoding T8: A Molecule Defining Functional Classes Of T Lympho- cytes", Cell, 40, pp. 237-46 (1985)]. Both predicted protein sequences define molecules with domains expected for surface antigens, including transmem-
30   brane and intracytoplasmic domains at the carboxyl end of the protein. In addition, both proteins con- tain an amino terminal region which shows striking homology to immunoglobulin and T cell receptor variable regions and which might function during
35   target cell recognition [Maddon et al., supra].

-3-

In immunocompetent individuals, T4 lymphocytes interact with other specialized cell types of the immune system to confer immunity to or defense against infection [E. L. Reinherz and S. F. Schlossman, "The Differentiation Function Of Human T-Cells", Cell, 19, pp. 821-27 (1980)]. More specifically, T4 lymphocytes stimulate production of growth factors which are critical to a functional immune system. For example, they act to stimulate B cells, the descendants of hemopoietic stem cells, which promote the production of defensive antibodies. They also activate macrophages ("killer cells") to attack infected or otherwise abnormal host cells and they induce monocytes ("scavenger cells") to encompass and destroy invading microbes.

It has been found that the primary target of or receptor for certain infective agents is the T4 surface protein. These agents include, for example, viruses and retroviruses. When T4 lymphocytes are exposed to such agents, they are rendered nonfunctional. As a result, the host's complex immune defense system is destroyed and the host becomes susceptible to a wide range of opportunistic infections.

Such immunosuppression is seen in patients suffering from acquired immune deficiency syndrome ("AIDS"). AIDS is a disease characterized by severe or, typically, complete immunosuppression and attendant host susceptibility to a wide range of opportunistic infections and malignancies. In some cases, AIDS infection is accompanied by central nervous system disorders. Complete clinical manifestation of AIDS is usually preceded by AIDS related complex ("ARC"), a syndrome accompanied by symptoms such as persistent generalized lymphadenopathy, fev r and weight loss. The human immunod - ficiency virus ("HIV") retrovirus is thought to be

-4-

the etiological agent responsible for AIDS infection
and its precursor, ARC [M. G. Sarngadharan et al.,
"Detection, Isolation And Continuous Production Of
Cytopathic Retroviruses (HTLV-III) From Patients
5   With AIDS And Pre-AIDS", Science, 224, pp. 497-508
(1984)].*

Between 85 and 100% of the AIDS/ARCS popu-
lation test seropositive for HIV [G. N. Shaw et al.,
"Molecular Characterization Of Human T-Cell Leukemia
10 (Lymphotropic) Virus Type III In The Acquired Immune
Deficiency Syndrome", Science, 226, pp. 1165-70
(1984)]. The number of adults in the United States
infected with HIV has been estimated to be between 1
and 2.5 million [D. Barnes, "Strategies For An AIDS
15 Vaccine", Science, 233, pp. 1149-53 (1986); M. Rees,
"The Sombre View Of AIDS", Nature, 326, pp. 343-45
(1987)]. These estimates include 64,900 individuals
who do not belong to an identified group at risk for
AIDS [S. L. Sivak and G. P. Wormser, "How Common Is
20 HTLV-III Infection In The United States?", New Eng.
J. Med., 313, p. 1352 (1985)]. The apparent annual
rate of diagnosis for those infected with HIV virus
is between 1 and 2% -- a rate which may increase
significantly in future years.

25   The genome of retroviruses, such as HIV,
contains three regions encoding structural proteins.
The gag region encodes the core proteins of the
virion. The pol region encodes the virion RNA-depen-
dent DNA polymerase (reverse transcriptase). The

30 _____

*   In this application, human immunodeficiency
virus ("HIV"), the generic term adopted by the human
retrovirus subcommittee of the International Committee
On Taxonomy Of Viruses to refer to independent iso-
35 lates from AIDS patients, including human T cell
lymphotropic virus type III ("HTLV-III"), lympha-
denopathy-associated virus ("LAV"), human immuno-
deficiency virus type 1 ("HIV-1") and AIDS-associated
retrovirus ("ARV") will be used.

env region encodes the major glycoprotein found in
the membrane envelope of the virus and in the cyto-
plasmic membrane of infected cells. The capacity of
the virus to attach to target cell receptors and to
5   cause fusion of cell membranes are two HIV proper-
ties controlled by the env gene. These properties
are believed to play a fundamental role in the patho-
genesis of the virus.

         HIV env proteins arise from a precursor
10   polypeptide that, in mature form, is cleaved into a
large heavily glycosylated exterior membrane protein
of about 481 amino acids -- gp120 -- and a smaller
transmembrane protein of about 345 amino acids which
may be glycosylated -- gp41 [L. Ratner et al.,
15   "Complete Nucleotide Sequence Of The AIDS Virus,
HTLV-III", Nature, 313, pp. 277-84 (1985)].

         The host range of the HIV virus is asso-
ciated with cells which bear the surface glycopro-
tein T4. Such cells include T4 lymphocytes and
20   brain cells [P. J. Maddon et al., "The T4 Gene Encodes
The AIDS Virus Receptor And Is Expressed In The
Immune System And The Brain", Cell, 47, pp. 333-48
(1986)]. Upon infection of a host by HIV virus, the
T4 lymphocytes are rendered non-functional. The
25   progression of AIDS/ARCS syndromes can be correlated
with the depletion of T4$^+$ lymphocytes, which display
the T4 surface glycoprotein. This T cell depletion,
with ensuing immunological compromise, may be attri-
butable to both recurrent cycles of infection and
30   lytic growth from cell-mediated spread of the virus.
In addition, clinical observations suggest that the
HIV virus is directly responsible for the central
nervous system disorders seen in many AIDS patients.
         The tropism of the HIV virus for T4$^+$ cells
35   is believed to be attributed to the role of the T4
cell surface glycoprotein as the membrane-anchor d
virus recept r. Because T4 behaves as the HIV virus

-6-

receptor, its extracellular sequence probably plays
a direct role in binding HIV. More specifically, it
is believed that HIV envelope selectively binds to
the T4 epitope(s), using this interaction to initiate
entry into the host cell [A. G. Dalgelish et al.,
"The CD4 (T4) Antigen Is An Essential Component Of
The Receptor For The AIDS Retrovirus", Nature, 312,
pp. 763-67 (1984); D. Klatzmann et al., "T-Lymphocyte
T4 Molecule Behaves As The Receptor For Human Retro-
virus LAV", Nature, 312, pp. 767-68 (1984)]. Accord-
ingly, cellular expression of T4 is believed to be
sufficient for HIV binding, with the T4 protein
serving as a receptor for the HIV virus.

The T4 tropism of the HIV virus has been
demonstrated in vitro. When HIV virus isolated from
AIDS patients is cultured together with T helper
lymphocytes preselected for surface T4, the lympho-
cytes are efficiently infected, display cytopathic
effects, including multinuclear syncytia formation
and are killed by lytic growth [D. Klatzmann et al.,
"Selective Tropism Of Lymphadenopathy Associated
Virus (LAV) For Helper-Inducer T Lymphocytes",
Science, 225, pp. 59-63 (1984); F. Wong-Staal and
R. C. Gallo, "Human T-Lymphotropic Retroviruses",
Nature, 317, pp. 395-403 (1985)]. It has been demon-
strated that a cloned cDNA version of human T4, when
expressed on the surface of transfected cells from
non-T cell lineages, including murine and fibroblas-
toid cells, endows those cells with the ability to
bind HIV [P. J. Maddon et al., "The T4 Gene Encodes
The AIDS Virus Receptor And Is Expressed In The Immune
System And The Brain", Cell, 47, pp. 333-48 (1986)].

During the course of HIV infection, the
host mounts both a humoral and a cellular immune
response t the virus. These responses include the
appearance of antibodies which bind to a number of
viral products and which exhibit neutralizing effect

-7-

or antibody dependent cellular cytotoxic functions
[M. Guroff-Robert et al., "HTLV-III-Neutralizing
Antibodies In Patients With AIDS And AIDS-Related
Complex", Nature, 316, pp. 72-74 (1985); D. D. F.
Barin et al., "Virus Envelope Protein Of HTLV-III
Represents Major Target Antigen For Antibodies In
AIDS Patients", Science, 228, pp. 1094-96 (1985);
A. H. Rook et al., "Sera From HTLV-III/LAV Antibody
Positive Individuals Mediate Antibody Dependent
Cellular Cytotoxicity Against HTLV-III/LAV Infected
T Cells", J. Immunol., 138, pp. 1064-68 (1987)].
Epitopes of the HIV envelope have been identified as
important determinants in eliciting a neutralizing
antibody response. And, determinants in antibody
dependent cellular cytotoxicity ("ADCC") activity
include HIV env and, possibly, gag epitopes.

In the absence to date of effective treat-
ments for AIDS, many efforts have centered on preven-
tion of the disease. Such preventative measures
include HIV antibody screening for all blood, organ
and semen donors and education of AIDS high-risk
groups regarding transmission of the disease.

Experimental or early-stage clinical treat-
ment of AIDS and ARCS conditions have included the
administration of antiviral drugs, such as HPA-23,
phosphonoformate, suramin, ribavirin, azidothymidine
("AZT") and dideoxycytidine, which apparently inter-
fere with replication of the virus through reverse
transcriptase inhibition. Although each of these
drugs exhibits activity against HIV in vitro, only
AZT has demonstrated potential benefits in clinical
trials. AZT administration in effective amounts,
however, has been accompanied by undesirable and
debilitating side effects, such as bone marrow
depression. It is likely, therefore, that hemato-
logic toxicity will b a major rate limiting factor
in the long term use of AZT.

Other proposed methods for treating AIDS
have focused on the development of agents having
activity against steps in the viral replicative cycle
other than reverse transcription. Such methods

5   include the administration of interferons or the
application of hybridoma technology. Most of these
treatment strategies are expected to require the
co-administration of immunomodulators, such as inter-
leukin-2.

10   To date, the need exists for the develop-
ment of effective immunotherapeutic agents and methods
for the treatment of AIDS, ARCS, HIV infection and
other immunodeficiencies caused by T lymphocyte
depletion or abnormalities.

15   ## DISCLOSURE OF THE INVENTION

The present invention solves the problems
referred to above by providing, in large amounts,
soluble T4 and soluble derivatives thereof that act
as receptors for infective agents whose primary target

20   is the T4 surface protein of $T4^+$ lymphocytes. Advan-
tageously, this invention also provides soluble T4
essentially free of other proteins of human origin
and in a form that is not contaminated by viruses,
such as HIV or hepatitis B virus.

25   As will be appreciated from the disclosure
to follow, the DNA sequences and recombinant DNA
molecules of this invention are capable of directing,
in an appropriate host, the production of soluble T4
or derivatives thereof. The polypeptides of this

30   invention are useful, either as produced in the host
or after further derivatization or modification, in
a variety of immunotherapeutic compositions and
methods for treating immunodeficient patients
suffering from diseases caused by infective agents

35   whose primary targets are $T4^+$ lymphocytes. According
to various embodiments of this invention, such compo-

-9-

sitions and methods relate to a soluble receptor for
HIV, soluble T4 proteins and polypeptides and anti-
bodies thereto. The soluble T4 proteins and polypep-
tides of this invention include monovalent, as well
5   as polyvalent forms.

The compositions and methods of this inven-
tion, which are based upon soluble T4 proteins, poly-
peptides or peptides and antibodies thereto, are
particularly useful for the prevention, treatment or
10 detection of the HIV-related infections AIDS and
ARC. More specifically, the soluble T4-based com-
positions and methods of this invention employ
soluble T4-like polypeptides -- polypeptides which
advantageously interfere with the T4/HIV interaction
15 by blocking or competitive binding mechanisms which
inhibit HIV infection of cells expressing the T4
surface protein. These soluble T4-like polypeptides
inhibit adhesion between $T4^+$ lymphocytes and infec-
tive agents which target $T4^+$ lymphocytes and inhibit
20 interaction between $T4^+$ lymphocytes and antigen pre-
senting cells and targets of $T4^+$ lymphocytes mediated
killing. By acting as soluble virus receptors, the
compositions of this invention may be used as anti-
viral therapeutics to inhibit HIV binding to $T4^+$
25 cells and virally induced syncytium formation at the
level of receptor binding.

This invention accomplishes these goals by
providing DNA sequences coding on expression in an
appropriate unicellular host for soluble T4 proteins[*]
30 and soluble derivatives thereof.

---

[*]   As used in this application, "soluble T4 pro-
tein", "soluble T4" and "soluble T4-like polypeptides"
include all proteins, polypeptides and peptides which
35 are natural or recombinant soluble T4 proteins, or
soluble derivatives thereof, and which are charac-
terized by the immunotherapeutic (anti-retroviral)

-10-

This invention also provides recombinant DNA molecules containing those DNA sequences and unicellular hosts transformed with them. Those hosts permit the production of large quantities of the

5 novel soluble T4 proteins, polypeptides, peptides and derivatives of this invention for use in a wide variety of therapeutic, prophylactic and diagnostic compositions and methods.

The DNA sequences of this invention are

10 selected from the group consisting of:

(a) the DNA inserts of p199-7, pBG377, pBG380, pBG381, p203-5, pBG391, pBG392, pBG393, pBG394, pBG395, pBG396, pBG397, p211-11, p214-10 and p215-7;

15 (b) DNA sequences which hybridize to one or more of the foregoing DNA inserts and which code on expression for a soluble T4-like polypeptide; and

(c) DNA sequences which code on expression for a soluble T4-like polypeptide coded for on

20 expression by any of the foregoing DNA inserts and sequences.

According to an alternate embodiment, this invention also relates to a DNA sequence comprising the DNA insert of p170-2, said sequence coding on

25 expression for a T4-like polypeptide. And, this invention also relates to recombinant DNA molecules and processes for producing T4 protein using that DNA sequence.

---

30 (footnote continued from preceding page)

or immunogenic activity of soluble T4 protein. They include soluble T4-like compounds from a variety of sources, such as soluble T4 protein derived from natural sources, recombinant soluble T4 protein and

35 synthetic or semi-synthtic soluble T4 protein.

-11-

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is an autoradiograph depicting the purification of T4 protein from U937 cells by immunoaffinity chromatography.

Figure 2 depicts autoradiograph and Western blot data demonstrating that immunoaffinity-purified, solubilized native T4 protein binds to HIV envelope protein.

Figure 3 depicts the nucleotide sequence and the derived amino acid sequence of T4 cDNA obtained from PBL clone λ203-4. In this figure, the amino acids are represented by single letter codes as follows:

Phe:	F	Leu:	L	Ile:	I	Met:	M
Val:	V	Ser:	S	Pro:	P	Thr:	T
Ala:	A	Tyr:	Y	His:	H	Gln:	Q
Asn:	N	Lys:	K	Asp:	D	Glu:	E
Cys:	C	Trp:	W	Arg:	R	Gly:	G

* = position at which a stop codon is present.

In Figure 3, the T4 protein translation start (AA$_{-23}$) is located at the methionine at nucleoides 201-203 and the mature N-terminus is located at the lysine (AA$_3$) at nucleotides 276-278.

Figure 4 is a schematic outline of the construction of cDNA clones pBG312.T4 (also called p171-1) and p170-2.

Figure 5 is a schematic outline of the construction of plasmid pEC100.

Figure 6 depicts amino acid comparisons at a positions 3, 64 and 231 of various T4 cDNA clones.

Figures 7A and 7B depict the protein domain structure of purified, solubilized T4 protein and recombinant soluble T4 mutants.

Figures 8A-8D are schematic outlines of constructions of various intermediate plasmids and other plasmids used to express recombinant soluble T4 ("rsT4") of this invention.

Figure 9A is a schematic outline of the construction of plasmid p199-7.

Figures 9B and 9C are schematic outlines of the construction of plasmid p203-5.

Figure 10 depicts the synthetic oligonucleotide linkers employed in various constructions according to this invention.

Figure 11 depicts the nucleotide sequence of the entire plasmid defined by p199-7 ($P_L$mutet.rsT4) and its rsT4.2 insert and the amino acid sequence deduced from the rsT4 sequence. This includes the ClaI-ClaI cassette which defines the Met perfect rsT4.2 coding sequence.

Figure 12 depicts a protein blot analysis of an induction of rsT4.2 expression from SG936/p199-7.

Figure 13 is a schematic outline of the construction of plasmid pBG368.

Figures 14A-14C are schematic outlines of constructions of various plasmids of this invention.

Figure 15 depicts the nucleotide sequence of plasmid pBG391.

Figure 16 depicts the nucleotide sequence of plasmid pBG392. In this figure, the T4 protein translation start (AA_{-23}) is located at the methionine at nucleotides 1207-1209 and the mature N-terminus is located at the lysine (AA_3) at nucleotide 1281-84.

Figure 17 is a schematic outline of constructions of various plasmids of this invention.

Figure 18 depicts the synthetic oligonucleotide linkers employed in various constructions according to this invention.

Figure 19 depicts the nucleotide sequence of plasmid pBG394.

Figure 20 depicts the nucleotide sequence of plasmid pBG396.

Figure 21 depicts the nucleotide sequence of plasmid pBG393.

Figure 22 depicts the nucleotide sequence of plasmid pBG395.

Figure 23 is a Coomassie stained gel of rsT4.2 purified from the conditioned medium of the pBG380 transfected CHO cell line BG380G of plasmid p196-10.

Figure 24 is a schematic outline of the construction of plasmid p196-10.

Figure 25 is a schematic outline of the construction of plasmid pBG394.

Figure 26 is a schematic outline of the construction of plasmid p211-11.

Figure 27 is a schematic outline of the construction of plasmid p215-7.

Figure 28 is a schematic outline of the construction of plasmid p218-8.

Figure 29A is a Coomassie stained gel of rsT4.113.1 purified from the conditioned medium of pBG211-11 transfected E.coli.

Figure 29B is an autoradiograph depicting a Western blot analysis of rsT4.113.1 expressed in E.coli.

Figure 30, panels (a)-(c) depict the purification of rsT4.113.1 from E.coli transformants.

Figure 31, panels (a)-(c) depict the refolding of purified rsT4.113.1.

Figure 32 is an autoradiograph depicting the immunoprecipitation of ^{35}S-metabolically labelled CHO cell lines producing recombinant soluble T4.

Figure 33 depicts an immunoblot analysis of COS 7 cell lines producing recombinant soluble T4.

Figure 34 depicts in graphic form the results of a competition assay between rsT4.113.1 and rsT4.3 for binding to OKT4A or OKT4.

Figures 35-37 depict in graphic form the results of competition assays between rsT4.111 and rsT4.3 for binding to, respectively, OKT4A, Leu-3A and OKT4.

Figure 38 depicts in graphic form an ELISA assay for rsT4.113.1 from *E.coli* transformants.

Figure 39 depicts in graphic form the results of a p24 radioimmunoassay using recombinant soluble T4 according to this invention.

Figures 40 and 41 depict the results of syncytia inhibition assays using recombinant soluble T4 proteins according to this invention.

Figure 42 is a schematic outline of the construction of plasmid pBiv.1.

Figure 43 depicts the bivalent recombinant soluble T4 protein produced by pBiv.1.

DETAILED DESCRIPTION OF THE INVENTION

We isolated the DNA sequences of this invention from two libraries: a λgt cDNA library derived the T cell tumor line REX and a λgt10 cDNA library derived from peripheral blood lymphocytes. However, we could also have employed libraries prepared from other cells that express T4. These include, for example, H9 and U937. We also used a human genomic bank to isolate various fragments of the T4 gene.

For screening these libraries, we used a series of chemically synthesized anti-sense oligonucleotide DNA probes based upon the T4 protein sequence set forth in Maddon et al. (1985), supra.

For screening, we hybridized our oligonucleotide probes to our cDNA libraries utilizing a plaque hybridization screening assay. We selected clones hybridizing to several of our probes. And, after isolating and subcloning the cDNA inserts of the sel cted clones int plasmids, w d termined

their nucleotide sequences and compared the amino
acid sequences deduced from those nucleotide sequences
to the amino acid sequences referred to in <u>Maddon</u>
<u>et al.</u> (1985), supra. As a result of these compari-
5 sons, we determined that all of our selected clones
were characterized by cDNA inserts coding for amino
acid sequences of human T4.

We have depicted in Figure 3 the nucleo-
tide sequence of full-length T4 cDNA obtained from
10 deposited clone p170-2 and the amino acid sequence
deduced therefrom. That cDNA sequence was subse-
quently subjected to <u>in vitro</u> site-directed mutagen-
esis and restriction fragment substitution so that
its cDNA sequence was identical to that of <u>Maddon</u>
15 <u>et al.</u>

After modifying our T4 cDNA sequence to be
identical to that of <u>Maddon et al.</u>, we truncated
samples of it in various positions to remove the
coding regions for the transmembrane and intracyto-
20 plasmic domains. The remaining cDNA sequences encoded
a soluble T4 which retained the extracellular region
believed to be responsible for HIV binding.

We then constructed various clones charac-
terized by such cDNA inserts coding for human soluble
25 T4. Those cDNA sequences may be used in a variety
of ways in accordance with this invention. More
particularly, those sequences or portions of them,
or synthetic or semi-synthetic copies of them, may
be used as DNA probes to screen other human or animal
30 cDNA or genomic libraries to select by hybridization
other DNA sequences that are related to soluble T4.
Typically, conventional hybridization conditions,
e.g., about 20° to 27°C below Tm, are employed in
such selections. However, less stringent conditions
35 may be necessary when the library is being screened
with a probe from a different species than that from

which the library is derived, e.g., the screening of a mouse library with a human probe.

Such cDNA inserts, portions of them, or synthetic or semi-synthetic copies of them, may also be used as starting materials to prepare various mutations. Such mutations may be either degenerate, i.e., the mutation does not change the amino acid sequence encoded by the mutated codon, or non-degenerate, i.e., the mutation changes the amino acid sequence encoded by the mutated codon. Both types of mutations may be advantageous in producing or using soluble T4's according to this invention. For example, these mutations may permit higher levels of production or easier purification of soluble T4 or higher T4 activity.

For all of these reasons, the DNA sequences of this invention are selected from the group consisting.of:

(a) the DNA inserts of p199-7, pBG377, pBG380, pBG381, p203-5, pBG391, pBG392, pBG393, pBG394, pBG395, pBG396, pBG397, p211-11, p214-10 and p215-7;

(b) DNA sequences which hybridize to one or more of the foregoing DNA inserts and which code on expression for a soluble T4-like polypeptide; and

(c) DNA sequences which code on expression for a soluble T4-like polypeptide coded for on expression by any of the foregoing DNA inserts and sequences.

Preferably, the DNA sequences of this invention code for a polypeptide selected from the group consisting of a polypeptide of the formula $AA_{-23}-AA_{362}$ of Figure 3, a polypeptide of t   formula AA_{1-362} of Figure 3, a polypeptide of the f: la Met-AA_{1-362} of Figure 3, a polypeptide of th   ɔrmula AA_{1-374} of Figure 3, a polypeptide of the fc   .a Met-AA_{1-374} of Figure 3, a polypeptide of tʰ   rmula AA_{1-377} of Figure 3, a polypeptide of the f   .a

-17-

Met-AA$_{1-377}$ of Figure 3, a polypeptide of the formula
AA$_{-23}$-AA$_{374}$ of Figure 3, a polypeptide of the formula
AA$_{-23}$-AA$_{377}$ of Figure 3, or portions thereof.

DNA sequences according to this invention
5   also preferably code for a polypeptide selected from
the group consisting of a polypeptide of the formula
AA$_{-23}$-AA$_{182}$ of Figure 16, a polypeptide of the
formula AA$_1$-AA$_{182}$ of Figure 16, a polypeptide of
the formula Met-AA$_{1-182}$ of Figure 16, a polypeptide
10   of the formula AA$_{-23}$-AA$_{182}$ of Figure 16, followed by
the amino acids asparagine-leucine-glutamine-histidine-
serine-leucine, a polypeptide of the formula
AA$_1$-AA$_{182}$ of Figure 16, followed by the amino acids
asparagine-leucine-glutamine-histidine-serine-leucine,
15   a polypeptide of the formula Met-AA$_{1-182}$ of Figure 16,
followed by the amino acids asparagine-leucine-
glutamine-histidine-serine-leucine, a polypeptide of
the formula AA$_{-23}$-AA$_{113}$ of Figure 16, a polypeptide
of the formula AA$_1$-AA$_{113}$ of Figure 16, a polypeptide
20   of the formula Met-AA$_{1-113}$ of Figure 16, a polypeptide
of the formula AA$_{-23}$-AA$_{111}$ of Figure 16, a polypeptide
of the formula AA$_1$-AA$_{111}$ of Figure 16, a polypeptide
of the formula Met-AA$_{1-111}$ of Figure 16, a polypep-
tide of the formula AA$_{-23}$-AA$_{131}$ of Figure 16, a poly-
25   peptide of the formula AA$_1$-AA$_{131}$ of Figure 16, a
polypeptide of the formula Met-AA$_{1-131}$ of Figure 16,
a polypeptide of the formula AA$_{-23}$-AA$_{145}$ of Figure 16,
a polypeptide of the formula AA$_1$-AA$_{145}$ of Figure 16,
a polypeptide of the formula Met-AA$_{1-145}$ of Figure 16,
30   a polypeptide of the formula AA$_{-23}$-AA$_{166}$ of Figure 16,
a polypeptide of the formula AA$_1$-AA$_{166}$ of Figure 16,
a polypeptide of the formula Met-AA$_{1-166}$ of Figure 16,
or portions thereof.

Additionally, DNA sequences of this inven-
35   tion code for a polypeptide selected from the group
consisting of a polypeptide of the formula AA$_{-23}$-AA$_{362}$
of matur T4 protein, a polypeptid of th formula

-16-

AA_{1-362} of mature T4 protein, a polypeptide of the formula Met-AA_{1-362} of mature T4 protein, a polypeptide of the formula AA_{1-374} of mature T4 protein, a polypeptide of the formula Met-AA_{1-374} of mature T4 protein, a polypeptide of the formula AA_{1-377} of mature T4 protein, a polypeptide of the formula Met-AA_{1-377} of mature T4 protein, a polypeptide of the formula AA_{-23}-AA_{374} of mature T4 protein, a polypeptide of the formula AA_{-23}-AA_{377} of mature T4 protein, or portions thereof.

DNA sequences according to this invention also code for a polypeptide selected from the group consisting of a polypeptide of the formula AA_{-23}-AA_{182} of mature T4 protein, a polypeptide of the formula AA_{1}-AA_{182} of mature T4 protein, a polypeptide of the formula Met-AA_{1-182} of mature T4 protein, a polypeptide of the formula AA_{-23}-AA_{182} of mature T4 protein, followed by the amino acids asparagine-leucine-glutamine-histidine-serine-leucine, a polypeptide of the formula AA_{1}-AA_{182} of mature T4 protein, followed by the amino acids asparagine-leucine-glutamine-histidine-serine-leucine, a polypeptide of the formula Met-AA_{1-182} of mature T4 protein, followed by the amino acids asparagine-leucine-glutamine-histidine-serine-leucine, a polypeptide of the formula AA_{-23}-AA_{113} of mature T4 protein, a polypeptide of the formula AA_{1}-AA_{113} of mature T4 protein, a polypeptide of the formula Met-AA_{1-113} of mature T4 protein, a polypeptide of the formula AA_{-23}-AA_{111} of mature T4 protein, a polypeptide of the formula AA_{1}-AA_{111} of mature T4 protein, a polypeptide of the formula Met-AA_{1-111} of mature T4 protein, a polypeptide of the formula AA_{-23}-AA_{131} of mature T4 protein, a polypeptide of the formula AA_{1}-AA_{131} of mature T4 protein, a polypeptide of the formula Met-AA_{1-131} of matur T4 protein, a polypeptid of the formula AA_{-23}-AA_{145} of matur T4 protein, a p lypeptide of

the formula AA_1-AA_{145} of mature T4 protein, a polypep-
tide of the formula Met-AA_{1-145} of mature T4 protein,
a polypeptide of the formula AA_{-23}-AA_{166} of mature
T4 protein, a polypeptide of the formula AA_1-AA_{166}
5   of mature T4 protein, a polypeptide of the formula
Met-AA_{1-166} of mature T4 protein, or portions thereof.

   The amino terminal amino acid of mature
T4 protein isolated from T cells begins at lysine,
the third amino acid of the sequence depicted in
10 Figure 16. Accordingly, soluble T4 proteins also
include polypeptides of the formula AA_3-AA_{377} of
Figure 16, or portions thereof. Such polypeptides
include polypeptides selected from the group consist-
ing of a polypeptide of the formula AA_3 to AA_{362} of
15 Figure 16, a polypeptide of the formula AA_3 to AA_{374}
of Figure 16, a polypeptide of the formula AA_3-AA_{182}
of Figure 16, a polypeptide of the formula AA_3-AA_{113}
of Figure 16, a polypeptide of the formula AA_3-AA_{131}
of Figure 16, a polypeptide of the formula AA_3-AA_{145}
20 of Figure 16, a polypeptide of the formula AA_3-AA_{166}
of Figure 16, and a polypeptide of the formula
AA_3-AA_{111} of Figure 16. Soluble T4 proteins also
include the above-recited polypeptides preceded by
an N-terminal methionine group.
25   Soluble T4 protein constructs according to
this invention may also be produced by truncating
the full length T4 protein sequence at various posi-
tions to remove the coding regions for the transmem-
brane and intracytoplasmic domains, while retaining
30 the extracellular region believed to be responsible
for HIV binding. More particularly, soluble T4
polypeptides may be produced by conventional tech-
niques of oligonucleotide directed mutagenesis;
restriction digestion, followed by insertion of
35 linkers; or chewing back full length T4 protein with
enzymes.

Alternatively, soluble T4 polypeptides may be chemically synthesized by conventional peptide synthesis techniques, such as solid phase synthesis [R. B. Merrifield, "Solid Phase Peptide Synthesis.
5 I. The Synthesis Of A Tetrapeptide", J. Am. Chem. Soc., 83, pp. 2149-54 (1963)].

The DNA sequences of this invention code for soluble proteins and derivatives that are believed to bind to Major Histocompatibility Complex antigens
10 and envelope glycoprotein of certain retroviruses, such as HIV. Preferably, they also inhibit syncytium formation, believed to be the mode of intracellular HIV virus spread. And, they may inhibit interaction between T4$^+$ lymphocytes and antigen-presenting cells
15 and targets of T4$^+$ cell mediated killing. Most preferably, they also inhibit adhesion between T4$^+$ lymphocytes and infective agents, such as the HIV virus, whose primary targets are T4$^+$ lymphocytes.

The DNA sequences of this invention are
20 also useful for producing soluble T4 or its derivatives coded for on expression by them in unicellular hosts transformed with those DNA sequences. As well known in the art, for expression of the DNA sequences of this invention, the DNA sequence should be opera-
25 tively linked to an expression control sequence in an appropriate expression vector and employed in that expression vector to transform an appropriate unicellular host.

Such operative linking of a DNA sequence
30 of this invention to an expression control sequence, of course, includes the provision of a translation start signal in the correct reading frame upstream of the DNA sequence. If the particular DNA sequence of this invention being expressed does not begin
35 with a methionine, th start signal will result in an additional amino acid -- methionine -- being located at the N-terminus of the product. While

-21-

such methionyl-containing product may be employed
directly in the compositions and methods of this
invention, it is usually more desirable to remove
the methionine before use. Methods are available in
5   the art to remove such N-terminal methionines from
polypeptides expressed with them. For example,
certain hosts and fermentation conditions permit
removal of substantially all of the N-terminal
methionine in vivo. Other hosts require in vitro
10 removal of the N-terminal methionine. However, such
in vivo and in vitro methods are well known in the
art.

    A wide variety of host/expression vector
combinations may be employed in expressing the DNA
15 sequences of this invention. Useful expression
vectors, for example, may consist of segments of
chromosomal, non-chromosomal and synthetic DNA
sequences, such as various known derivatives of SV40
and known bacterial plasmids, e.g., plasmids from
20 E.coli including col E1, pCR1, pBR322, pMB9 and their
derivatives, wider host range plasmids, e.g., RP4,
phage DNAs, e.g., the numerous derivatives of phage λ,
e.g., NM989, and other DNA phages, e.g., M13 and
filamenteous single stranded DNA phages, yeast plas-
25 mids, such as the 2μ plasmid or derivatives thereof,
and vectors derived from combinations of plasmids
and phage DNAs, such as plasmids which have been
modified to employ phage DNA or other expression
control sequences. For animal cell expression, we
30 prefer to use plasmid pBG368, a derivative of pBG312
[R. Cate et al., "Isolation Of The Bovine And Human
Genes For Mullerian Inhibiting Substance And
Expression Of The Human Gene In Animal Cells", Cell,
45, pp. 685-98 (1986)] which contains the major late
35 promoter of adenovirus 2.

    In addition, any of a wide variety f
expression control sequenc s -- sequenc s that con-

trol the expression of a DNA sequence when opera-
tively linked to it -- may be used in these vectors
to express the DNA sequence of this invention. Such
useful expression control sequences, include, for

5   example, the early and late promoters of SV40 or the
adenovirus, the lac system, the trp system, the TAC
or TRC system, the major operator and promoter regions
of phage λ, the control regions of fd coat protein,
the promoter for 3-phosphoglycerate kinase or other

10   glycolytic enzymes, the promoters of acid phosphatase,
e.g., Pho5, the promoters of the yeast α-mating
factors, the polyhedron promoter of the baculovirus
system and other sequences known to control the
expression of genes of prokaryotic or eukaryotic

15   cells or their viruses, and various combinations
thereof. For animal cell expression, we prefer to
use an expression control sequence derived from the
major late promoter of adenovirus 2.

      A wide variety of unicellular host cells

20   are also useful in expressing the DNA sequences of
this invention. These hosts may include well known
eukaryotic and prokaryotic hosts, such as strains of
E.coli, Pseudomonas, Bacillus, Streptomyces, fungi,
such as yeasts, and animal cells, such as CHO and

25   mouse cells, African green monkey cells, such as
COS 1, COS 7, BSC 1, BSC 40, and BMT 10, insect cells,
and human cells and plant cells in tissue culture.
For animal cell expression, we prefer CHO cells and
COS 7 cells.

30   It should of course be understood that not
all vectors and expression control sequences will
function equally well to express the DNA sequences
of this invention. Neither will all hosts function
equally well with the same expression system. How-

35   ever, one of skill in the art may make a selection
among these vectors, expression control sequences,
and hosts without undue experimentation and without

departing from the scope of this invention. For
example, in selecting a vector, the host must be
considered because the vector must replicate in it.
The vector's copy number, the ability to control
5   that copy number, and the expression of any other
proteins encoded by the vector, such as antibiotic
markers, should also be considered.

In selecting an expression control sequence,
a variety of factors should also be considered.
10   These include, for example, the relative strength of
the system, its controllability, and its compatibility
with the particular DNA sequence of this invention,
particularly as regards potential secondary struc-
tures. Unicellular hosts should be selected by
15   consideration of their compatibility with the chosen
vector, the toxicity of the product coded for on
expression by the DNA sequences of this invention to
them, their secretion characteristics, their ability
to fold proteins correctly, their fermentation re-
20   quirements, and the ease of purification of the
products coded on expression by the DNA sequences of
this invention.

Within these parameters, one of skill in
the art may select various vector/expression control
25   system/host combinations that will express the DNA
sequences of this invention on fermentation or in large
scale animal culture, e.g., CHO cells or COS 7 cells.

The polypeptides produced on expression of
the DNA sequences of this invention may be isolated
30   from the fermentation or animal cell cultures and
purified using any of a variety of conventional
methods. One of skill in the art may select the
most appropriate isolation and purification tech-
niques without departing from the scope of this
35   invention.

The polypeptides produced on expression of
the DNA sequences of this invention are essentially

-24-

free of other proteins of human origin. Thus, they are different than T4 protein purified from human lymphocytes.

The polypeptides of this invention are

5 useful in immunotherapeutic compositions and methods. For example, the polypeptides of this invention are active in inhibiting infection by agents whose primary targets are T4$^+$ lymphocytes by interfering with their interaction with those target lymphocytes. More

10 preferably, the polypeptides of this invention may be employed to saturate the T4 receptor sites of T4-targeted infective agents. Thus, they exert anti-viral activity by competitive binding with cell surface T4 receptor sites. This effect is plainly

15 of great utility in diseases, such as AIDS, ARC and HIV infection. Accordingly, the polypeptides and methods of this invention may be used to treat humans having AIDS, ARC, HIV infection or antibodies to HIV. In addition, these polypeptides and methods

20 may be used for treating AIDS-like diseases caused by retroviruses, such as simian immunodeficiency viruses, in mammals, including humans.

According to one embodiment of this invention, antibodies to soluble T4 proteins and polypep-

25 tides may be used in the treatment, prevention, or diagnosis of AIDS, ARC and HIV infection.

The polypeptides of this invention may also be used in combination with other therapeutics used in the treatment of AIDS, ARC and HIV infection.

30 For example, soluble T4 polypeptides may be used in combination with anti-retroviral agents that block reverse transcriptase, such as AZT, HPA-23, phos-phonoformate, suramin, ribavirin and dideoxyciti-dine. Additionally, these polypeptides may be used

35 with anti-viral agents such as interferons, includ-ing alpha interferon, beta interferon and gamma interferon, or glucosidase inhibitors, such as

-25-

castanospermine. Such combination therapies advantageously utilize lower dosages of those agents, thus avoiding possible toxicity.

And, the polypeptides of this invention may be used in plasmapheresis techniques or in blood bags for selective removal of viral contaminants from blood. According to this embodiment of the invention, soluble T4 polypeptides may be coupled to a solid support, comprising, for example, plastic or glass beads, or a filter, which is incorporated into a plasmapheresis unit.

Additionally, the compositions of this invention may be employed as immunosuppressants useful in preventing or treating graft-vs-host disease, autoimmune diseases and allograft rejection.

The compositions of this invention typically comprise an immunotherapeutic effective amount of a polypeptide of this invention and a pharmaceutically acceptable carrier. Therapeutic methods of this invention comprise the step of treating patients in a pharmaceutically acceptable manner with those compositions.

The compositions of this invention for use in these therapies may be in a variety of forms. These include, for example, solid, semi-solid and liquid dosage forms, such as tablets, pills, powders, liquid solutions or suspensions, liposomes, suppositories, injectable and infusable solutions. The preferred form depends on the intended mode of administration and therapeutic application. The compositions also preferably include conventional pharmaceutically acceptable carriers and adjuvants which are known to those of skill in the art.

Generally, the pharmaceutical compositions of the present invention may be formulated and administered using methods and compositions similar to those used for other pharmaceutically important poly-

peptides (e.g., alpha-interferon). Thus, the poly-
peptides may be stored in lyophilized form, reconsti-
tuted with sterile water just prior to administration,
and administered by the usual routes of administration
5   such as parenteral, subcutaneous, intravenous, intra-
muscular or intralesional routes. An effective dosage
may be in the range of from 0.5 to 5.0 mg/kg body
weight/day, it being recognized that lower and higher
doses may also be useful.

10   This invention also relates to soluble
receptors and their use in diagnosing or treating
viral agents which target or bind to those receptors.
Such soluble receptors may be used as decoys to
absorb viral agents and to halt the spread of viral
15   infection. Alternatively, virus-killing agents may
be attached to the soluble protein receptors,
providing a direct mode of delivery of those agents
to the virus.

         More particularly, the polypeptides of
20   this invention are useful in diagnostic compositions
and methods to detect or monitor the course of HIV
infection. Advantageously, these polypeptides are
useful in diagnosing variants of the HIV virus,
regardless of origin of the infecting HIV agent.

25   For example, soluble T4 proteins and poly-
peptides according to this invention, which have a
high affinity for HIV, may be advantageously used to
increase the sensitivity of HIV assay systems now
based upon monoclonal or polyclonal antibodies.
30   More specifically, soluble T4 proteins and polypep-
tides may be used to pretreat test plasma to concen-
trate any HIV present, even in small amounts, so
that it is more easily recognized by the antibody.
And soluble T4 proteins and polypeptides may be used
35   to purify the HIV envelope protein gp120.

         Alternatively, th soluble T4 proteins and
polypeptides of this invention may be used t replace

anti-HIV antibodies now used in various assays.
These soluble T4 proteins and polypeptides are be
preferable to anti-HIV antibodies for two reasons.
First, soluble T4, exhibits an affinity for HIV of
approximately 10^{-9}, a level which exceeds the 10^{-7}
to 10^{-8} values of anti-HIV antibodies. And, while
anti-HIV antibodies are more likely to be specific
for different HIV isolates, strain variations would
not affect a soluble T4 protein-based assay, since
all HIV isolates must be capable of interacting with
the T4 receptor as a prerequisite to infectivity.

For example, a soluble T4 protein or poly-
peptide may be linked to an indicator, such as an
enzyme, and used in an ELISA assay. Here, soluble
T4 advantageously acts as a measure of both HIV in a
test sample and any free HIV envelope gp120 protein.

And, polyvalent forms of soluble T4 proteins
or polypeptides may be produced, for example, by
chemical coupling or genetic fusion techniques, thus
increasing even further the avidity of soluble T4
for HIV.

In order that this invention may be better
understood, the following examples are set forth.
These examples are for purposes of illustration only,
and are not to be construed as limiting the scope of
the invention in any manner.

EXAMPLES

Purification Of Native Solubilized T4

We purified native T4 from the T4[+]-promono-
cytic cell line U937 derived from a histocytic
lymphoma to approximately 50% purity using immuno-
affinity chromatography as follows.

We grew U937 cells [a gift from Dr. Scott
Hammer, New England Deaconess Hospital] to
10^{6} cells/ml in RPMI 1640, 10% FCS, harv sted and
washed them in 1X PBS. We then lys d th cell pellet

in 20 mM Tris-HCl (pH 7.7), 0.5% NF-40 (a non-ionic detergent), 0.2% NaDOC, 0.2 mM EGTA, 0.2 mM PMSF and 5 µg/ml BPTI at 4×10^7 cells/ml. Because this purification was carried out in the presence of a

5   non-ionic detergent, T4, which is normally membrane-bound via its hydrophobic transmembrane domain, was isolated as a solubilized protein. We spun the lysate in a GS 3 rotor for 10 min at 10,000 rpm and stored the supernatant at -70°C.

10   Subsequently, we preabsorbed the clarified cell extract with mouse IgG-Sepharose, followed by protein A Sepharose and then passed the flowthrough through an immunoaffinity column comprising immobilized 19Thy anti-T4 monoclonal antibody on Affigel-10

15   [a gift from Dr. Ellis Reinherz, Dana Farber Cancer Institute, Boston, Massachusetts]. We washed the column extensively and eluted the bound material with 50 mM glycine-HCl (pH 2.5), 0.15 M NaCl, 0.5% NP-40, 5 µg/ml BPTI and 0.2 mM EGTA.

20   We then separated 10 µl aliquots of each elution fraction on a 10% SDS-PAGE under reducing conditions, with the bands being visualized by silver staining. As shown in Figure 1, a major silver-stained band of 55 Kd was visible. We then carried

25   out two assays on the 55 Kd protein and sequenced the amino terminus of the protein to confirm its identity as native solubilized T4.

Sequencing Of Native Solubilized T4

We determined the N-terminal amino acid
30   sequence of our solubilized native T4 which we isolated from a detergent extract of U937 cells by immunoaffinity chromatography as described above.

Techniques for determining the amino acid sequences of various proteins and peptides derived
35   from them are well known in the art. We chose automated Edman degradation to determine th amino

terminus of our solubilized native T4. More speci-
fically, we gel purified and electroeluted approxi-
mately 5 µg of the solubilized native T4 and then
subjected it to automated Edman degradation using a
5   gas phase sequencer (Applied Biosystems 470A). We
then identified the PTH-amino acids produced at each
cycle of the Edman chemistry by high pressure liquid
chromatography, on-line with the sequencer, in a
PTH-amino acid analyzer (Applied Biosystems 120A).
10   Direct analysis of the protein provided amino terminal
sequence information which, when compared to the
amino acid sequence deduced from the cDNA sequence
of human T4 [Maddon et al. (1985), supra], identified
the purified protein as human T4.

15   Radioimmunoassay Of Native Solubilized T4

To determine that our purification process
enriched for T4, we assayed fractions from the
immunoaffinity elution step in a T4-specific sandwich
radioimmunoassay, based upon the ELISA assay of P. E.
20   Rao et al., in Cellular Immunology, 80, pp. 310-19
(1983). We coated each well of a Removawell strip
(Dynatech Labs, Alexandria, Virginia) with 50 µl of
10 µl/ml OKT4 antibody (ATCC #CRL 8002) or MOPC195
(a background binding control) in 0.05 M sodium
25   bicarbonate buffer (pH 9.4) at 4°C overnight. We
washed the wells and then filled them with 1% FCS in
PBS to saturate the protein binding capacity of the
plastic. After removing the 1% FCS solution, we
added test samples, in 50 µl aliquots, to the wells.
30   We then incubated the samples for 4 hours at room
temperature. Subsequently, we removed the samples
and washed the wells four times with 0.05% Tween-20
in PBS. We then added ^{125}I-labelled 19Thy antibody
(50,000-100,000 cpm per well) and incubated the wells
35   at 4°C overnight. We then washed the wells four

times and separated each well for bound ^{125}I detection
in a Beckman gamma detector.

5   As shown in Figure 1, in which values were
plotted following subtraction for background, the
peak fraction of solubilized native T4 protein
detected by radioimmunoassay coincided with elution
of the 55 Kd protein seen by silver staining.

Western Blot Assay For T4

          Although many antibodies have been developed
10   for detecting T4 antigen, none are useful for protein
blot analysis (Dr. Ellis Reinherz, personal communi-
cation). In order to develop antibodies useful for
Western blot detection of soluble T4 to follow the
purification of T4 and recombinant soluble T4, we
15   raised polyclonal, hyperimmune anti-T4 antisera in
rabbits against three synthetic T4 oligopeptides.
These oligopeptides are represented in Figure 3 as
follows:

Oligopeptide	Amino Acid Coordinates
JB-1	44-63
JB-2	133-156
JB-3	325-343

We had previously synthesized these peptides using
conventional phosphoamide DNA synthesis techniques.
25   See, e.g., Tetrahedron Letters, 22, pp. 1859-62
(1981). We synthesized the peptides on an Applied
Biosystems 380A DNA Synthesizer and purified them by
gel electrophoresis.

(i) Coupling Of T4 Peptides To BTG

30   We coupled each of these peptides to the
carrier protein bovine thyrogobulin ("BTG") [Sigma,
St. Louis, Missouri] according to a modification of
procedures set forth in J. Rothbard et al., J. Exp.
Med., 160, pp. 208-21 (1984) and R. C. Kennedy et al.,
35   "Antiserum To A Synthetic Peptide Rec gnizes Th

HTLV-III Envelope Glycoprotein", Science, 231, pp. 1556-59 (1986).

More specifically, we mixed 10 mg of BTG diluted in 1 ml of PBS with 1.3 mg of m-maleimido-benzoyl-N-hydroxysuccinimide ester ("MBS") in 0.5 ml of dimethylformamide ("DMF"). We mixed the reaction mixture well and reacted it for about 1 hour at 25°C. Subsequently, we loaded the mixture onto a Sephadex G25 gel filtration column (Pharmacia, Sweden) which had been pre-equilibrated with 0.1 M PBS (pH 6.0). We then collected a total of thirty 2 ml aliquot elution fractions and read the absorbance of each fraction at 280 nm ("A_{280}"). We then pooled the three peak fractions (15, 16 and 17) to create the activated carrier.

We dissolved 10 mg of $NaBH_4$ in 2.5 ml of 0.1 M sodium borate solution to produce a sodium borohydride solution. Subsequently, we diluted approximately 8 mg of each of synthetic T4 peptides JB-1, JB-2 and JB-3 with 1 ml of 0.1 M borate buffer and then mixed each solution with 200 µl of the sodium borohydride solution, incubating the mixture on ice for 5 minutes. We then warmed each peptide solution to 25°C, brought each solution to pH 1.0 with 1 N HCl (during which frothing occurred) and then brought each solution to pH 7.0 with 1 N NaOH (after the frothing had stopped).

We then coupled each peptide to BTG by adding 1.2 ml of the peptide solution to 6 ml of the activated carrier solution. We allowed the coupling reaction to proceed overnight by incubating the reaction mixture at room temperature.

(ii) Inoculation Of Test Animals

We dissolved each of the BTG-coupled peptides prepared above in sterile Freund's complete adjuvant, to a final c ncentration of 1 µg/ml coupled

peptide in PBS. Subsequently, we inoculated each of
three rabbits (New Zealand white) by intramuscular
injection of 500 µg of one of the coupled peptides
into each rabbit. We inoculated a fourth rabbit
5   (New Zealand white) in the same manner with a mixture
of the three coupled peptides. All rabbits were
prebled prior to boosting to establish an average
baseline for each response to be measured. The
rabbits were boosted at 6 weeks with 500 µg coupled
10 peptide in incomplete Freund's adjuvant.

        Serum was collected from each rabbit monthly
for 4 months after immunization. The serum was then
assayed for antipeptide titer.

15   (iii) **ELISA With Antipeptide Sera**
               Against Peptide Coated Plates

        In this assay, we determined that antiserum
raised in an animal by each of peptides JB-1, JB-2
and JB-3 binds to that peptide. Accordingly, those
peptides are immunogenic and elicit a response in
20 test animals.

        To carry out the assay, we coated Immulon-2
(Dynatech Labs, Alexandria, Virginia) microtiter
plates with 50 µl per well of 50 µg/ml uncoupled
peptide in PBS and incubated the plates overnight at
25 4°C. Plates coated with peptide 46R*, which served
as controls, were treated identically. We then washed
the plates 4 times with PBS-Tween (0.5%) and 4 times
with water. The plates were blotted dry by gentle
tapping over paper towels. After blotting the plates,

30 _____

        *   Peptide 46 corresponds to amino acids ("AA")
728-751 of the _env_ gene of the HIV genome. The amino
acid numbering corresponds to that set forth for the
env gene in L. Ratner et al., "Complete Nucleotide
35 S quence Of Th AIDS Virus, HTLV-III", _Nature_, 313,
pp. 277-84 (1985). Peptide 46 has the sequence:
                LPIPRGPDRPEGIEEEGGERDRDR.

-33-

we added 200 µl of a 5% FCS/PBS solution to each
well and incubated the plates for 1 hour at room
temperature.

     We then assayed serum samples from the
5 rabbits on the pre-coated plates prepared as described
above. We assayed the antibody response to the
immunogen peptide at an initial dilution of 1:100,
followed by serial 10-fold dilutions in 5% FCS/PBS.

     After a 2 hour incubation period at room
10 temperature, we washed the plates and blotted them
dry as described above. We then added 50 µl of a
1:1500 dilution of horseradish peroxidase ("HRP")-
conjugated goat anti-rabbit-IgG [Cooper Biomedical,
Malvern, Pennsylvania] in 5% FCS/PBS to each well
15 and incubated the plates at room temperature for
1 hour. We washed the plates with PBS-Tween 0.5%.
We then added 50 µl of 0.42 mM TMB. We stopped the
enzyme reactions with 50 µl of 2 M H_2SO_4. We then
analyzed the plates spectrophotometrically at 450 nm
20 using a microtiter plate reader [Dynatech Labs,
Alexandria, Virginia].

     We observed that antiserum against each of
peptides JB-1, JB-2 and JB-3 binds to the corre-
sponding peptide. We also observed that antiserum
25 against a mixture of peptides JB-1, JB-2 and JB-3
binds to peptides JB-1 and JB-3 under the conditions
set forth above. The titers of each of the four
antisera tested against the peptides in the solid-
phase ELISA are shown below, where "ND" represents
30 values not determined:

	Approximate Titer Against:		
Peptide	JB-1	JB-2	JB-3
JB-1	>1/50,000	0	ND
JB-2	0	1/50,000	ND
JB-3	0	0	1/10,000
JB-1 + JB-2 + JB-3	1/4,000	ND	1/7,000

Ig fractions from two of the three anti-
peptide sera raised against individual peptides,
anti-JB-1 and anti-JB-2, recognized the 55 Kd T4
antigen band of native solubilized T4 in a Western

5   blot analysis of protein eluted from the 19Thy
(anti-T4) monoclonal antibody affinity column
described above. As in the case of the radioimmuno-
assay of native solubilized T4, the detection of the
55 Kd protein coincides with its apparent elution

10   from the affinity column. This provides further
evidence that our T4 purification procedure enriched
for solubilized T4.

Thus, these polyclonal sera are useful in
the detection of nanogram quantities of T4 (both

15   native and recombinant forms) by Western analysis.

Binding of Cell-Free T4 To HIV Envelope

We then tested our purified solubilized
native T4 isolated from U937 cells for its ability
to bind to the HIV envelope protein gp160/gp120. To

20   carry out this direct binding assay, we incubated
^{35}S-labelled gp160/gp120 detergent cell extract
derived from a recombinant cell line 7d2 (a gift
from Drs. Mark Kowalski and William Haseltine, Dana-
Farber Cancer Institute) with samples of solubilized

25   native T4, each of which had been preincubated with
one type of monoclonal antibody.

More specifically, we mixed 5 µl of solu-
bilized T4 in a microfuge tube with 5 µg (about 3 µl)
of OKT4 (ATCC #CRL 8002), a monoclonal antibody

30   recognizing an epitope on T4 which does not interfere
with HIV binding [J. A. Hoxie et al., J. Immunol.,
136, pp. 361-63 (1986)] or with 5 µg of OKT4A (Ortho
Diagnostics #7142), a monoclonal antibody that inter-
feres with HIV binding to T4 positive cells [J. Steven

35   McDougal et al., J. Immunol., 137, pp. 2937-2944
(1986)]. Alternativ ly, we mixed 50 µl of solubilized

-35-

T4 with 5 μg of αHTLV III gp120 (Dupont #NEN-9284).
We then incubated the mixtures on ice for 1 hour.

Subsequently, we added 150 μl of ^{35}S-
labelled gp160/gp120 cell extract or ^{35}S-labelled

5   control cell extract (precleared with protein-A
Sepharose) to the preincubated solubilized T4/mono-
clonal antibody mixtures and rocked the tubes over-
night at 4°C. We then precipitated the T4/gp160/gp120
immune complexes by adding 30 μl of protein-A

10   Sepharose to each tube and rocking for 2 hours at
4°C to allow the protein-A Sepharose to bind to the
antibody complexes. Subsequently, we spun down the
beads in an Eppendorf microfuge and after extensive
washings, we eluted with 40 μl SDS sample buffer at

15   65°C for 10 minutes. We then loaded 20 μl of the
eluted material on a 7.5% SDS-PAGE gel which was run
under reducing conditions.

Figure 2 depicts autoradiograph and Western
blot results of the T4/gp160/gp120 coimmunoprecipita-

20   tions. In Figure 2, lanes 1-5 were autoradiographed
after treatment with 40% sodium salicylate and lanes
6-7 were developed on a Western blot with rabbit
antisera JB-2.

As shown in Figure 2, gp160/gp120 protein

25   was coimmunoprecipitated in the presence of T4 with
OKT4 (lane 5) but not in the presence of T4 with
OKT4A (lane 4). Lane 3 shows the positive control
for gp160/gp120 using αHTLV III gp120 monoclonal
antibody. Neither negative control with ^{35}S-labelled

30   control extract (lane 1) or protein-A Sepharose alone
(lane 2) showed bands migrating in the position of
gp160/gp120. Based upon the bands that developed
on the Western blot, the amount of T4 precipitated
with either OKT4 (lane 6) or OKT4A (lane 7) appeared

35   to be similar.

This demonstrates that purified, solubilized
nativ T4, which is naturally membrane bound, can

still interact with the HIV glycoprotein in solution.
Accordingly, we believe that cell free soluble T4 is
useful in preventing the binding interaction between
HIV and the T4 receptor of T4$^+$ lymphocytes. By com-
5   peting with cell surface T4 for binding to the HIV
envelope protein gp120, soluble T4 is useful in block-
ing HIV infection.

Synthesis Of Oligonucleotide DNA Probes

The nucleotide sequence and a deduced amino
10   acid sequence for a cDNA that purportedly encodes
the entire human T4 protein have been reported
[Maddon et al., (1985), supra]. The deduced primary
structure of the T4 protein reveals that it can be
divided into domains as demonstrated below:

15

Structure/Proposed Location	Amino Acid Coordinates
Hydrophobic/Secretory Signal	-23 to -1
Homology to V-Regions/ Extracellular	+1 to +94
Homology to J-Regions/ Extracellular	+95 to +109
Glycosylated Region/ Extracellular	+110 to +374
Hydrophobic/Transmembrane Sequence	+375 to +395
Very Hydrophilic/ Intracytoplasmic	+396 to +435

Based on the sequence for the above-listed
domains, we chemically synthesized antisense
30   oligonucleotide DNA probes using conventional phos-
phoamide DNA synthesis techniques. See, e.g.,
Tetrahedron Letters, 22, pp. 1859-62 (1981). We
synthesized the probes on an Applied Biosystems 380A
DNA synth sizer and purified them by gel electro-
35   phoresis.

-37-

Furthermore, we synthesized the probes
such that they were complementary to the DNA
sequences which code for the amino acid sequence,
i.e., the probes were antisense, to enable them to
5   recognize and hybridize to the corresponding sequences
in DNA, as well as in mRNA. The nucleotide sequences
of the eleven selected regions of the T4 protein
[corresponding to the nucleotide numbering set forth
in <u>Maddon et al.</u>, (1985), supra] were the following:

Oligonucleotide	Nucleotide Coordinates
1	145-171
2	742-765
3	1414-1440
6	427-453
7	1303-1329
8	1012-1038
9	97-118
10	10-36
11	1698-1724
12	397-423
14	261-287

Before using our DNA probes for screening,
we 5' end-labelled each of the single-stranded DNA
25 probes with ^{32}P using $[\gamma-^{32}P]$-ATP and T4 polynucleo-
tide kinase, substantially as described by A. M. Maxam
and W. Gilbert, "A New Method For Sequencing DNA",
<u>Proc. Natl. Acad. Sci. USA</u>, 74, pp. 560-64 (1977).

Construction of λgt10 Peripheral Blood
30 <u>Lymphocytes cDNA Library</u>

To prepare our Peripheral Blo d Lymphocytes
(PBL) cDNA library, w pr cessed PBL, from a single
leukophoresis donor, through one round of absorption

to remove monocytes. We then stimulated the non-
adherent cells with IFN-γ 1000 U/ml and 10 μg/ml PHA
for 24 hours. We isolated RNA from these cells using
phenol extraction [Maniatis et al., Molecular Cloning,
5   p. 187 (Cold Spring Harbor Laboratory) (1982)] and
prepared poly A⁺ mRNA by one round of oligo dT cellu-
lose chromatography. We ethanol precipitated the
RNA, dried it in a speed vac and resuspended the RNA
in 10 μl H_2O (0.5 μg/μl). We treated the RNA for 10
10   min at room temperature in CH_3HgOH (5 mM final con-
centration) and β-mercaptoethanol (0.26 M). We then
added the methyl mercury treated RNA to 0.1 M Tris-HCl
(pH 8.3) at 43°C, 0.01 M Mg, 0.01 M DTT, 2 mM Vanadyl
complex, 5 μg oligo dT_{12-18}, 20 mM KCl, 1 mM dCTP,
15   dGTP, dTTP, 0.5 mM dATP, 2 μCi[α-^{32}P]dATP and 30 U
1.5 μl AMV reverse transcriptase (Seikagaku America)
in a total volume of 50 μl. We incubated the mixture
for 3 minutes at room temperature and then for 3 hours
at 44°C, after which time we stopped the reaction by
20   the addition of 2.5 μl of 0.5 M EDTA.

        We extracted the reaction mixture with an
equal volume of phenol:chloroform (1:1) and precipi-
tated the aqueous layer two times with 0.2 volume of
10 M NH_4AC and 2.5 volumes EtOH and dried it under
25   vacuum. The yield of cDNA was 1.5 μg.

        We synthesized the second strand according
to the methods of Okayama and Berg [Mol. Cell. Biol.,
2, p. 161 (1982)] and Gubler and Hoffman [Gene, 25,
pp. 263-69 (1983)], except that we used the DNA poly-
30   merase I large fragment in the synthesis.

        We blunt ended the double-stranded cDNA by
resuspending the DNA in 80 μl TA buffer (0.033 M Tris
Acetate (pH 7.8); 0.066 M KAcetate; 0.01 M MgAcetate;
0.001M DTT; 50 μg/ml BSA), 5 μg RNase A, 4 units RNase
35   H, 50 μM β NAD , 8 units E.coli ligase, 0.3125 mM
dATP, dCTP, dGTP, and dTTP, 12 units T_4 polymerase
and incubated the reacti n mixture for 90 min at

37°C, added 1/20 volume of 0.5M EDTA, and extracted
with phenol:chloroform. We chromatographed the
aqueous layer on a G150 Sephadex column in 0.01M
Tris-HCl (pH 7.5), 0.1 M NaCl, 0.001 M EDTA and

5 collected the lead peak containing the double-stranded
cDNA and ethanol precipitated it. Yield: 0.605 μg
cDNA.

We ligated the double-stranded cDNA to
linker 35/36:

10
5'AATTCGAGCTCGAGCGCGGCCGC3'

3' GCTCGAGCTCGCGCCGGCG5'

using standard procedures. We then size selected
the cDNA for 800 bp and longer fragments on a S500
Sephacryl column, and ligated it to EcoRI-digested

15 bacteriophage lambda vector gt10 (a gift of
Dr. Ellis Reinherz). We packaged aliquots of the
ligation reaction in Gigapak (Strategene) according
to the manufacturer's protocol. We used the packaged
phage to infect E.coli BNN102 cells and plated the

20 cells for amplification. The resulting library con-
tained 1.125×10^6 independent recombinants.

We also screened a PBL cDNA library in the
bacteriophage lambda vector gt10 (a gift of Dr. Ellis
Reinherz), which was synthesized from mRNA from a

25 T4$^+$ tumor cell line named REX, which expresses T4
protein at high levels [O. Acuto et al., "The Human
T Cell Receptor: Appearance In Ontogeny And
Biochemical Relationship Of Lambda and Beta Subunits
on IL-2 Dependent Clones And T Cell Tumors", Cell,

30 34, pp. 717-26 (1983)].

Screening Of The Libraries

We then used three of our ^{32}P-labelled
synthetic oligonucleotide antisense probes, probes 3,
6 and 9, to screen in parallel our two λgt10 cDNA

35 libraries using th plaque hybridization screening
technique described in R. Cate et al., "Isolation Of

-40-

The Bovine And Human Genes For Mullerian Inhibiting
Substance And Expression Of The Human Gene In Animal
Cells", Cell, 45, pp. 685-98 (1986), with minor
modifications. We modified the Cate et al. proce-
5   dure by hybridizing without tetramethyl ammonium
chloride to accommodate our use of unique probes,
rather than mixtures, to probe the plaque filters.

We used the three probes, which had been
previously 5' end-labelled with $[\gamma-^{32}P]$-ATP according
10   to the method of A. Maxam and W. Gilbert, Meth.
Enzymol., 68, pp. 499-80 (1979) to screen in parallel
the PBL cDNA library and the REX cDNA library dis-
cussed above.

From our screening of the PBL library, we
15   isolated a nearly full length soluble T4 cDNA clone --
λ203-4 (or λgt10.PBL.T4) -- containing a 3.064 kb
insert which could be cleaved from the λgt10 vector
with EcoRI.

From our screening of the REX cell library,
20   we isolated an incomplete T4 cDNA clone containing
a 1,200 bp cDNA insert. We then further characterized
the DNA from these clones by DNA sequencing analysis.

We also screened a bacteriophage lambda
human genomic library, constructed in the vector
25   EMBL3 by Dr. Mark Pasek (Biogen Inc., Cambridge,
Massachusetts) [N. Murray in Lambda 2, eds. R. Hendrix,
J. Roberts, F. Stahl, R. Weisberg, pp. 3935-422 (1983)].
The library contains DNA fragments, created by partial
restriction of chromosomal DNA from the human lympho-
30   blastid cell line GM1416,48, XXXX (Human Genetic
Mutant Cell Repository, Camden, New Jersey) with
Sau3a, ligated onto EMBL3 arms which had been sub-
jected to cleavage with BamHI according to the pro-
cedures outlined in Maniatis et al., (1982), supra.
35   Plating of the phag library, lysis, and transfer of
the phag DNA ont nitrocellulose wer performed as
describ d by W. D. Bent n and R. W. David, "Screening

-41-

of Lambda gt Recombinant Clones By Hybridization To
Single Plaques In Situ", Science. 196, p. 180 (1977)
and Maniatis et al. (1982). Hybridization conditions
were those described by Cate et al. (1986), supra,
5   except that tetramethylammonium chloride (TMACl) was
omitted from the washing buffer.

Approximately 2 million plaques were
screened in parallel hybridizations with probe 1
and probe 3 discussed above. One phage, called
10 CM47, which hybridized with probe 3 in the primary
screenings, was subjected to DNA sequence analysis
to determine the existence and position of an intron
between the coding sequences for the predicted extra-
cellular and transmembrane domains. No phage clones
15 containing T4 sequences were found screening with
probe 1, probably because it includes a sequence
interrupted by an intron [D. R. Littman and S. N.
Gettner, Nature, 325, pp. 453-55 (1987); and our
observations].

20   Partial sequence analysis of CM47 shows
that an intron interrupts the sequence corresponding
to the codon for valine (amino acid 363) of the
deduced primary sequence for T4 (Figure 3 -- in which
introns are indicated by a solid line). This intron
25 defines a potential site for introducing a stop codon
in order to express a soluble form of T4. Another
intron found within the coding sequence for T4 inter-
rupts the codon for arginine (amino acid 295) and a
third intron in CM47 is found between the codons for
30 arginine (amino acid 402) and arginine (amino acid
403) (Figure 3).

Sequencing Of cDNA Clones

We then subcloned EcoRI digested DNA from
clone λ203-4 into animal expression vect r pBG312
35 [R. Cate et al., supra] to facilitate sequence
analysis. More specifically, as depicted in Figure 4,

we then digested λgt10.PBL.T4 with EcoRI to excise
the 3.064 kbp EcoRI-EcoRI fragment containing the
full length T4 cDNA. This cDNA sequence, including
the entire coding region for soluble T4 and for full
5   length T4 was deposited in p170-2. We used T4 ligase
to ligate the fragment into animal expression vector
pBG312 [supra] which had been previously cut with
EcoRI, to form pBG312.T4 and p170-2 (Figure 4). We
then determined the nucleotide sequence of the EcoRI
10   fragment of pBG312.T4 using Maxam Gilbert technology
[A. M. Maxam and W. Gilbert, "A New Method For
Sequencing DNA", Proc. Natl. Acad. Sci. USA, 74,
pp. 560-64 (1977)] (see Figure 3, which depicts the
PBL cDNA sequence in comparison to that reported by
15   Maddon et al., (1985), supra). This analysis showed
that the 3.064 kbp PBL full length complementary DNA
copy of T4 cDNA contained the coding sequence for
T4, approximately 200 bp of 5' noncoding sequence
and approximately 1500 bp of 3' noncoding sequence.
20   We then cut pBG312.T4 with PstI and removed
the resulting 3' protruding ends with Klenow and
isolated an approximately 2.5 kbp fragment. We then
inserted the fragment into the polylinker of pBG312
(which had been previously restricted at the SmaI
25   site) to form plasmid p170-2, which contains the
full length PBL T4 cDNA sequence (see Figure 3).
          As depicted in Figure 3, the PBL T4 cDNA
contains a nucleotide sequence almost identical to
the approximately 1,700 bp sequence reported by
30   Maddon et al., (1985), supra. The PBL T4 cDNA, how-
ever, contains three nucleotide substitutions that,
in the translation product of this cDNA, would pro-
duce a protein containing three amino acid substi-
tutions compared to the sequence reported by Maddon
35   et al. As shown in Figure 3, these differenc s are
at amino acid position 3, wh re the asparagin of
Maddon et al. is replac d with lysine; position 64,

where the tryptophan of Maddon et al. is replaced
with arginine and at position 231, where the phenyl-
alanine of Maddon et al. is replaced with serine.
The asparagine reported at position 3 of Maddon
et al. instead of lysine was the result of a sequenc-
ing error (Dr. Richard Axel; personal communication).
The significance of the amino acid replacements at
positions 64 and 231, which may represent allellic
polymorphism [T. C. Fuller et al., Human Immunology,
9, pp. 89-102 (1982); W. Stohl and H. G. Kunkel,
Scand. J. Immunol., 20, pp. 273-78 (1984); N. Amino
et al., Lancet, 2, pp. 94-95 (1984); and M. Sato
et al., J. Immunol., 132, pp. 1071-73 (1984)], is
not known.

DNA sequence analysis [Maxam and Gilbert,
supra] of the insert in pEC100 of the REX clone sug-
gests that it represents the product of a splicing
error, because 5' noncoding sequence appears to have
.been spliced with coding sequence beginning with the
GGT codon for glycine (amino acid 49) (see Figure 3
and Figure 5). The T4 coding sequence in pEC100*
from glycine (amino acid 49) to isoleucine (amino
acid 435) is identical to the sequence of Maddon
et al., (1985), supra.

In comparison, our earlier N-terminal pro-
tein sequence analysis of native T4 protein purified
from U937 cells shows a T4 expression product with
aspargine as amino acid 3. These differences are
also set forth in Figure 6, which also depicts com-
parisons at corresponding positions of the partial
clone from the REX cell line λgt10 library; our

---

* We constructed pEC100 by digesting the incomplete
T4 cDNA clone from the REX library with EcoRI and
isolating the 1,200 bp cDNA insert. We then ligated
it t pUC12 (Boehring r Mannheim, Indianapolis,
Indiana) which had been pr viously cut with EcoRI to
form pEC100.

-44-

genomic clone from a λEMBL3 library; mouse T4
sequences [Tourvieille et al., Science, 234, p. 61C
(1986)] and sheep T4 sequences [Classon et al.,
Immunogenetics, 23, p. 129 (1986)].

5   Construction of Soluble T4 Mutants

           We then employed the technique of in vitro
site-directed mutagenesis and restriction fragment
substitution to modify the T4 cDNA coding sequence
of p170-2 in sequential steps to be identical to
10   that reported by Maddon et al., (1985), supra. We
first used oligonucleotide-directed mutagenesis to
modify the amino acids at positions 3 and 64. Next,
we employed restriction fragment substitution with a
fragment including the serine 231 codon of a partial
15   T4 cDNA isolated from a T4 positive lymphocyte cell
line [O. Acuto et al., Cell, 34, pp. 717-26 (1983)]
library in λgt11 (a gift from Dr. Ellis Reinherz),
to modify the amino acid at position 231. We then
truncated our modified T4 cDNA sequence to remove
20   the coding regions for the transmembrane and intra-
cytoplasmic domains. Subsequently, we constructed
three different soluble T4 mutants from our full
length T4 clone PBL T4 by linker insertion between
restriction sites in order to increase the probability
25   of empirically finding a stable, secretable T4 mole-
cule. The structure of each of these mutants is
depicted in Figure 7A.

           Line A of Figure 7A represents a hydropathy
analysis of our full length soluble T4 carried out
30   using a computer program called Pepplot (University
of Wisconsin Genetics Computer Group) according to
J. Kyte and R. F. Doolittle, J. Mol. Biol., 157,
pp. 105-32 (1982). Line B depicts the protein domain
structure of full length T4 [Maddon et al., (1985)
35   supra] in which "S" repres nts the secretory signal
sequence, "V" represents the immunoglobulin-lik

variable region sequence, "J" represents the immuno-
globulin-like joining region sequence, "U" represents
the unique, extracellular region sequence, "TM"
represents the transmembrane sequence and "C" repre-
5   sents the cytoplasmic region sequence. In line B,
the transmembrane amino acid sequence and some flank-
ing sequence is written below the TM domain. Line C
depicts the protein domain structure of recombinant
soluble T4 mutants rsT4.1 in pBG377, rsT4.2 in pBG380
10   and rsT4.3 in pBG381. Line D represents the protein
domain structure of E.coli rsT4 gene (Met-perfect
construct) (p199-7) which is deleted for the T4
N-terminal signal sequence (S).

            We constructed the first three soluble T4
15   mutant gene fragments by truncating our full length
soluble T4 cDNA at positions corresponding to either
intron/exon boundaries or to protein domain boundaries
defined by hydropathy analysis predictions. More
specifically, we introduced synthetic linkers into
20   the unique AvaI site that is 5' to the transmembrane/
extracellular domain boundary to produce an in-frame
translational stop codon, thus constructing T4 genes
that lack the transmembrane and cytoplasmic domains
of the full length T4 sequence.

25   For example, mutant rsT4.1 in pBG377 was
truncated by the insertion of a stop codon following
amino acid 362, lysine, which corresponds to the
position of an intron separating the extracellular
and transmembrane domain exons. The positions both
30   of this intron and of the adjacent intron that splits
the transmembrane and cytoplasmic domains were deter-
mined by DNA sequence analysis of chromosomal T4
clones isolated from the λEMBL3 genomic library
described above. Although the significance of the
35   intron positions flanking the T4 transmembrane domain
is not known, the determinati n of the genetic struc-
ture c uld provide important information for design-

ing rsT4 mutants, since exons frequently define functional domains [W. Gilbert, "Why Genes In Pieces?", Nature, 271, p. 501 (1978)].

We then constructed mutant rsT4.2 in pBG380
5   by truncating the T4 cDNA at the boundary of the transmembrane and extracellular domains at amino acid 374. And, we constructed mutant rsT4.3 in pBG381 by truncating the T4 cDNA at amino acid 377, three amino acids downstream from the transmembrane/
10   extracellular domain boundary and within the trans- · membrane domain.

We also employed the technique of oligo-nucleotide site directed mutagenesis, according to D. Strauss et al., "Active Site Of Triosephosphate
15   Isomerase: In Vitro Mutagenesis And Characterization Of An Altered Enzyme", Proc. Natl. Acad. Sci. USA, 82, pp. 2272-76 (1985), to construct a fourth soluble T4 mutant from our full length T4 clone PBL T4. The structure of this mutant is depicted in Figure 7A,
20   line D, which represents the protein domain structure of E.coli rsT4 gene (Met-perfect rsT4.2) construct, deposited in p199-7, which is deleted for the T4 N-terminal signal sequence (S).

We also constructed various other soluble
25   T4 deletion mutants to determine which smaller fragments of the T4 sequence provide a protein which binds to HIV. These constructions were based on our belief that only the amino terminal sequence of T4 is required for binding to HIV. This belief, in turn,
30   was based upon observations that the monoclonal anti-body OKT4A blocks infection of T4 positive cells by HIV and it appears to recognize an epitope in the amino portion of T4 [Fuller et al., supra]. Such fragments of T4, which lack glycosylation and which
35   are capabl of binding HIV and blocking infection, may be produc d in E.coli or chemically synthesized.

-47-

The structure of each of these deletion mutants is depicted in Figure 7B. In that figure, line A depicts the protein domain structure of full length T4 [Maddon et al., (1985), supra; Figure 7A].

5   In line B, the protein structure of recombinant soluble T4 mutants are depicted as follows: rsT4.7 in p203-5, rsT4.7 in pBG392, rsT4.8 in pBG393, rsT4.9 in pBG394, rsT4.10 in pBG395, rsT4.11 in pBG397, rsT4.12 in pBG396, rsT4.111 in pBG215-7, rsT4.113.1

10   in pBG211-11 and rsT4.113.2 in pBG214-10.

We constructed soluble T4 derivatives p203-5, pBG392, pBG393, pBG394 and pBG396 by truncating our rsT4.2 gene after the StuI sites at amino acids 183 and 264 of rsT4.2. More specifically, we

15   constructed derivative rsT4.7 in p203-5 and in pBG392 by truncating the rsT4.2 cDNA at amino acid 182. And, we constructed each of derivatives rsT4.9 in pBG394 and rsT4.12 in pBG396 by truncating the rsT4.2 cDNA at amino acids 113, and 166, respectively. One

20   may also construct each of derivatives rsT4.10 in pBG395 and rsT4.11 in pBG397 by truncating the rsT4.2 cDNA at amino acids 131 and 145, respectively.

Expression of T4 and Soluble T4 Polypeptides In Bacterial Cells

25   The cDNA sequences of this invention can be used to transform eukaryotic and prokaryotic host cells by techniques well known in the art to produce recombinant soluble T4 polypeptides in clinically and commercially useful amounts.

30   For example, we constructed expression vector p199-7, as shown in Figure 9A, as follows.

We preceded the construction depicted in Figure 9A by the construction of various intermediate plasmids, as depicted in Figures 8A-8D. Those con-

35   structions w re carried out using conventional

recombinant techniques. The linkers employed in those constructions are set forth in Figure 10.

As depicted in Figures 8A and 8B, starting with p170-2, which contains our full length T4 DNA sequence, coding for T4 characterized by three different amino acids than that of <u>Maddon et al.</u>, (1985), supra, we produced various constructs which direct the expression of soluble T4. Some of these constructs are characterized in that one or more of those amino acid differences have been changed to correspond to the respective amino acids of <u>Maddon et al.</u> In this figure, as well as in the other figures, amino acid changes are reflected by an arrow.

Plasmid p192-6 contains the Met perfect rsT4.2 sequence derived by oligonucleotide site-directed mutagenesis which removed the entire T4 N-terminal signal sequence as shown in Figure 8C. And, to provide a convenient means of transferring the rsT4.2 Met perfect sequence into <u>E.coli</u> expression vectors, the steps described in Figure 8D were carried out to produce p195-8, a plasmid containing the Met perfect rsT4.2 sequence flanked by <u>Cla</u>I restriction sites. The <u>Cla</u>I-<u>Cla</u>I cassette of p195-8 optimizes the distance between the 5' <u>Cla</u>I site and the initiating Met codon. In Figure 8D, ST8 rop⁻ is a tetracycline resistance encoding pAT153-based plasmid containing the rop⁻ mutation that permits high plasmid copy number, a promoter and ribosome binding site from bacteriophage gene 32 and the gene 32 transcription termination sequence.

Cleavage of p195-8 with <u>Cla</u>I produced the fragment used to assemble p199-7, a construction which directs the expression of Met perfect rsT4.2 under the control of the P_L promoter (Figure 9A). As the first step, to construct a vect r from which rsT4.2 expression is under contr l of th P_L promot r,

-49-

we constructed the vector p197-12 from p1034
(p1muGCSF) (Figure 9A).

We then cut p1034 with EcoRI and BamHI to
excise the GCSF cDNA insert and a portion of the

5   phage mu ribosome binding site sequence -- which we
subsequently reconstructed with oligonucleotides.
The synthetic linkers used were linkers 57-60
(Figure 10).

We then ligated the synthetic linker into

10   the EcoRI/BamHI-cut p1034 to form p197-12. One could,
instead, replace these steps by starting with any
suitable E.coli expression vector containing a ClaI
site appropriately placed between the promoter and
terminator sequences. We cut p197-12 with ClaI and

15   inserted a ClaI-ClaI cassette containing the cDNA
sequence of rsT4.3 in pBG381 and phage transcription
terminator derived from p1034. The sequence of this
cassette is depicted in Figure 11. The resulting
plasmid, p199-7, contains the rsT4.2 "Met perfect"

20   gene in that vector.

Alternatively, one could derive the Met
perfect rsT4.2 sequence from plasmid pBG380,
deposited in connection with this application, and
gap out the signal sequence to create p192-6.

25   We tested for expression of p199-7 as
follows. SG936, an E.coli lon htpr double mutant
[ATCC 39624] [S. Goff and A. Goldberg, "ATP-Dependent
Protein Degradation In E.coli", in Maximizing Gene
Expression, W. Reznikoff and L. Gold (eds.) (1986)],

30   was transformed with p199-7 by conventional proce-
dures [Maniatis et al. (1982)] to form SG936/p199-7,
a transformant containing a plasmid with the Met-
perfect rsT4.2 gene behind the P_L promoter. Trans-
formants were selected on LB agar plates containing

35   10 mcg/ml tetracycline (t t). After str aking out
several single colonies for single colony isolation,
on was chosen at random for testing induction of

rsT4.2 synthesis. We picked a single colony from an
LB-agar tet$^+$ plate into 20 ml Luria Broth (LB) and
10 mcg/ml tet in a 125 ml shake flask and grew it
overnight in a shaking air incubator (New Brunswick
5   Scientific, New Jersey) at 30°C.

We then initiated an induction culture by
adding 0.5 ml of the overnight culture to 50 ml LB
and tet in a 500 ml flask which was grown at 30°C in
a shaking air incubator. When the culture reached
10   an OD(600) of 0.4, we transferred it to a 42°C water-
bath and shook it gently for approximately 20 minutes.
After heat induction at 42°C, the flask was trans-
ferred to a 39°C air incubator (New Brunswick
Scientific, New Jersey) where it was shaken vigorously
15   at 250 rpm. We withdrew samples just after the 42°C
heat shock, and at hourly time points for 4 hours,
and then after overnight growth. The samples were
measured for growth by OD(600) and analyzed following
SDS-PAGE for the pattern of protein synthesis by
20   Coomassie blue protein staining and by Western blot
analysis with our rabbit antipeptide antibody probes
(described above). Based on the relative molecular
weight and protein blot analysis, the expression of
rsT4.2 was induced from SG936/p199-7 following heat
25   induction at 42°C (Figure 12).

We transformed p199-7 into a $P_L$mu.tet
expression vector, an E.coli expression vector, at
the unique ClaI site (see Figure 11). The nucleo-
tide and amino acid sequences of p199-7 are shown in
30   Figure 11.

The expression of soluble T4 from p199-7
in E.coli was measured by Western blot analysis of
whole cell extracts following SDS-PAGE using the
rabbit polyclonal anti-peptide JB-1 or anti-peptide
35   JB-2 antibodies as probes (Figure 12).

We also constructed xpression v ctor
p203-5, as shown in Figure 9B, as follows.

We started with p197-7, which has the same
sequence as the P$_L$um vector p197-12 (see Figure 9A),
except that there is a single nucleotide deletion in
the 5' noncoding region following the P$_L$ promoter.
5   That deletion, which is a deletion of nucleotide
#40 -- adenine -- of p197-12 (see Figure 11), resulted
from a deletion in the region that was constructed
from linkers 57-60 (see Figure 10). p197-7 contains
the rsT4.2 gene comprising 374 amino acids. Alter-
10   natively, one could also use p197-7 as a starting
plasmid.

We cut p197-7 with ClaI. We also cut p195-8
(see Figures 8D and 9A) with ClaI to remove the
ClaI - ClaI cassette containing the cDNA sequence of
15   rsT4.2. Subsequently, we inserted the ClaI-ClaI
cassette into p197-7 to produce p198-2.

We then digested p198-2 with StuI to
remove 80 amino acids (amino acid 185 to amino acid
264) of the mature T4 protein coding sequence. Unex-
20   pected methylation, however, prevented cutting at
the second StuI site, so that only the StuI site at
amino acid 184 was cleaved. Following ligation, the
plasmid DNA was transformed into E.coli and we
examined several plasmid clones for the deletion
25   using standard procedures. None of those plasmids
contained the expected StuI deletion.

Subsequent DNA sequence analysis of one
of these plasmids, called p203-5, showed that two
guanine residues (see amino acids 183 and 184;
30   nucleotides 818 and 819 of Figure 3) of the StuI
recognition sequence had been deleted following
cleavage due to exonuclease digestion caused by the
use of exonuclease-contaminated StuI enzyme. This
dinucleotide deletion produced a translation frame-
35   shift following amino acid 182 (glutamine) and intro-
duced a stop codon six amino acid codons downstream
from the frameshift (Figure 9C). The unexpected

methylation of the second StuI site together with
the deletion that resulted in a new stop codon
produced a gene encoding a shortened form of recom-
binant soluble T4, called rsT4.7. The rsT4.7 sequence

5   encodes a 182 amino acid N-terminal segment of the
mature T4 sequence followed by, at the C-terminus,
six amino acids -- asparagine-leucine-glutamine-
histidine-serine-leucine -- of non-T4 sequence and
finally by a TAA stop codon.

10   The expression of soluble T4 from p203-5
in E.coli was measured by Western blot analysis
as previously described.

Expression of T4 and Soluble T4
Polypeptides In Animal Cells

15   We inserted both soluble T4 genes and the
unmodified gene encoding membrane-bound T4 into
animal expression vector pBG368. More specifically,
we inserted each of the soluble gene constructs into
pBG368 under the transcriptional control of the

20   adenovirus late promoter, to give plasmids pBG377,
pBG380 and pBG381. We also made two pBG312-based
constructions, called pBG378 and pBG379, which
direct the expression of recombinant full length T4
protein. pBG378 and pBG379 code for the same full

25   length T4 protein but in pBG379, a portion of the 3'
untranslated sequence has been removed. Subsequently,
to test for expression of recombinant soluble T4 and
recombinant full length T4, we cotransfected Chinese
hamster ovary ("CHO") cells with one of each of

30   those plasmids and with the plasmid pAdD26.
We first constructed pBG368 as follows.
As depicted in Figure 13, we cut animal cell expres-
sion vector pBG312 [R. Cate et al., "Isolation Of
The Bovine And Human Genes For Mullerian Inhibiting

35   Substance And Expression Of The Human Gen In Animal
Cells", Cell, 45, pp. 685-98 (1986)] with EcoRI and

-53-

BglII to delete one of each of the two EcoRI and the
two BglII restriction sites (the EcoRI site at posi-
tion 0 and the BglII site located at approximately
position 99). The resulting plasmid, pBG368, retained
an EcoRI site in the cloning region and a BglII site
after the cloning region. This left a single EcoRI
site and a single BglII site in the polylinker for
cloning purposes.

More specifically, we deleted one EcoRI
site and one BglII site by sequential partial diges-
tion of pBG312 with restriction enzymes EcoRI and
BglII, respectively. We filled in with Klenow and 4
nucleotides then religated to produce pBG368, which
contains unique restriction sites for EcoRI and BglII
enzymes.

Once transient expression of soluble T4
was verified, we constructed stable cell lines that
continuously expressed soluble T4. To do this, we
employed the stable cell expression host, the
dihydrofolate reductase deletion mutant (DHFR⁻)
Chinese hamster ovary cell line [F. Kao et al.,
"Genetics Of Somatic Mammalian Cells X Complementation
Analysis of Glycine-Requiring Mutants", Proc. Natl.
Acad. Sci., 64, pp. 1284-91 (1969); L. Chasin and
G. Urlab "Isolation Of Chinese Hamster Cell Mutants
Deficient In Dihydrofolate Reductase Activity",
Proc. Natl. Acad. Sci., 77, pp. 4216-80 (1980)].

Using this system, we cotransfected each
T4 gene construct with pAdD26 [R. J. Kaufman and
P. A. Sharp, "Amplification And Expression Of
Sequences Cotransfected With a Modular Dihydrofolate
Reductase Complementary DNA Gene", J. Mol. Biol.,
159, pp. 661-21 (1982) containing the mouse DHFR
gene. Before carrying out the co-transfections, we
linearized all plasmids by restriction enzym cleavage
and, prior to transfection, we mixed each plasmid
with pAdD26 so that the molar rati of pAdD26 to T4

-54-

was 1:10. This maximized the number of T4 gene copies
per transfectant.

Within the cell, the plasmids were ligated
together to form polymers that can become integrated
into host chromosomal sequences by illegitimate
recombination [J. Haynes and C. Weissmann, "Constitu-
tive, Long-Term Production Of Human Interferons By
Hamster Cells Containing Multiple Copies Of a Cloned
Interferon Gene", Nucl. Acids Res., 11, pp. 687-706
(1983); S. J. Scahill et al., "Expression And Charac-
terization Of The Product Of A Human Immune Interferon
cDNA Gene In Chinese Hamster Ovary Cells", Proc. Natl.
Acad. Sci. USA, 80, pp. 4654-58 (1983)]. We selected
transfectants that express the mouse DHFR gene in
culture medium lacking nucleotides. We then subjected
these transfectants to a series of increasing concen-
trations of methotrexate, a toxic folate analogue
that binds DHFR, to select for cells levels of DHFR.

Resistance to methotrexate by increased
expression of DHFR is frequently the result of DHFR
gene amplification, which can include the reiteration
of large chromosomal segments, called amplified
units [R. J. Kaufman and P. A. Sharp, "Amplification
And Expression Of Loss Of Dihydrofolate Reductase
Genes In A Chinese Hamster Ovary Cell Line", Molec.
Cell. Biol., 1, pp. 1069-76 (1981)]. Therefore,
cointegration of DHFR and rsT4 sequences permitted
the amplification of rsT4 genes. Stably transfected
cell lines were isolated by cloning in selective
growth medium, then screened for T4 expression with
a T4 antigen (RIA) [D. Klatzmann et al., Nature,
312, pp. 767-68 (1984)] and by immunoprecipitation
from conditioned medium after [^{35}S] cysteine
("^{35}S-Cys") metabolic labelling.

We also inserted the solubl T4 derivative
rsT4.7 gene into an animal cell expression plasmid
as f llows.

-55-

As set forth in Figure 14C, we cut plasmid pBG381 (Figure 14A) with EcoRI and NheI. We then cut p186-6 with EcoRI and NheI to remove the 786 base pair fragment. We ligated that fragment into the

5   digested pBG381 to form plasmid pBG391. The T4 sequence in pBG391 is identical to both that of Maddon et al. (1985) supra at positions 64 (tryptophan) and 231 (phenylalanine) and to that of pBG381. However, at position 3, the asparagine reported by

10   Maddon et al. and present in pBG381 is replaced with lysine. The nucleotide sequence of pBG391 is depicted in Figure 15.

We then digested p203-5 with NheI and OxanI to remove the 483 base pair fragment. We

15   inserted that fragment into NheI/OxanI-digested pBG391 to form plasmid pBG392, the animal cell expression construct of rsT4.7. The T4 sequence in rsT4.7 contains amino acids identical to that of Maddon et al.'s full length sequence at amino acid

20   positions 64 (tryptophan) and 231 (phenylalanine). However, at position 3, the asparagine reported by Maddon et al. is replaced with lysine. The nucleotide sequence of pBG392 is depicted in Figure 16.

In Figure 14D, we have depicted the con-

25   struction of other animal cell expression constructs containing sequences encoding the deletions rsT4.9 in pBG394, and rsT4.12 in pBG396. Those constructions were carried out using conventional recombinant techniques. The linkers employed in those constructions

30   are set forth in Figure 18. The nucleotide sequences of pBG394 and pBG396 are shown in Figures 19 and 20.

Plasmid pBG393, shown in Figure 17, contains rsT4.8, the perfect form of rsT4.7. pBG393 contains 182 amino acids of the mature T4 sequence,

35   without the additional non-T4 6 amin acids at the C-terminus following amin acid 182. The nucleotide s quence of BG393 is shown in Figure 21.

Other animal cell expression plasmids
according to this invention may be constructed as
depicted in Figure 17. These include rsT4.10 in
pBG395 and rsT4.11 in pBG397 (see Figure 18 for
5   specific linkers).

The nucleotide sequence of BG395 is shown
in Figure 22.

Purification Of Recombinant Soluble T4

Recombinant soluble T4 construct pBG380
10 expressed in DHFR⁻ CHO cells was grown to confluency
in a α-Modified Eagles Medium (Gibco) supplemented
with 10% fetal calf serum, 1 mM glutamine and the
antibiotics penicillin and streptomycin (100 µg/ml
of each). The cells were grown at 37°C in two 21 Cell
15 Factory Systems (Nunc). We then washed the confluent
cells free of fetal calf serum with α-Modified Eagles
Medium without fetal calf serum and cultured the
cells in α-Modified Eagles Medium at 37°C for 4 days.
Subsequently, we harvested the conditioned media,
20 filtered it through a Millipore Millidisk 0.22µ
hydrophilic filter cartridge (Millipore #MCGL 305-01)
and concentrated the secreted proteins on a fast-S
ion exchange column (S-Sepharose Fast Flow, Pharmacia
#17-0511-01) in 20 mM MES buffer (pH 5.5).
25   We then eluted the bound proteins with 20 mM
Tris-HCl (pH 7.7) and 0.3 M NaCl. The elution pool
was subsequently diluted with 2 volumes of 20 mM
Tris-HCl (pH 7.7) and it was then loaded on a column
comprising immobilized 19Thy anti-T4 monoclonal anti-
30 body coupled to Affigel-10 [a gift of Dr. Ellis
Reinherz, Dana Farber Cancer Institute, Boston,
Massachusetts]. We washed the column extensively
and eluted the bound material as 0.5 ml fractions
with 50 mM glycine-HCl (pH 2.5), 150 mM NaCl, 0.1 mM
35 EGTA and 5 µg/ml bovine pancreatic trypsin inhibitor,
Apr tinin (Sigma #A1153). We used Western blots

-57-

developed with rabbit antisera raised against peptide
JB-2 to follow the purification. We employed silver
stained gels to follow binding and elution of rsT4.2
during the chromatography. Figure 23 depicts a
5   Coomassie stained gel of purified rsT4.2.

Gel sizing-column chromatography analysis
of the purified rsT4.2 from the pBG380 transfected
CHO cell line, BG380G, suggests that rsT4 is monmeric
under physiologic pH and salt concentration.

10   Sequencing Of Recombinant
Soluble T4 Protein

We then determined the N-terminal amino
acid sequence of a recombinant soluble T4, specifi-
cally rsT4.2, molecule purified from the conditioned
15   medium of the pBG380 transfected CHO cell line BG80G,
as described above, by automated Edman degradation
in an Applied Biosystems 470A gas phase sequenator
[R. B. Pepinsky et al., J. Biol Chem., 261,
pp. 4239-46 (1986)].

20   The amino terminal sequence matched the
sequence which we had previously determined for
solubilized native T4 isolated from U937 cells, supra.
The amino terminal sequences of native solubilized
T4 (sT4) and purified rsT4 protein are Δ2 proteins,
25   as compared to the amino terminal sequence predicted
by Maddon et al., (1985), supra, with the mature amino
terminus located at position 3 of that sequence. The
amino terminal sequences of solubilized native T4
(sT4), recombinant soluble T4 (rsT4.2) secreted by
30   CHO transfectant BG380G containing pBG380 and the
protein sequence deduced by Maddon et al. (1985),
supra are as follows:

sT4:   X-K-V-V-L-X-K-K-X-D-T-V-E-L-T-X-T-A-S-E-

rsT4.2:   N-K-V-V-L-G-K-K-G-D-T-V-E-L-T-X-T-A-S-E-

Maddon
et al.   Q-G-N-K-V-V-L-G-K-K-G-D-T-V-E-L-T-C-T-A-S-E

In the above sequences, the amino acids
are represented by single letter codes as follows:

5
Phe:	F	Leu:	L	Ile:	I	Met:	M
Val:	V	Ser:	S	Pro:	P	Thr:	T
Ala:	A	Tyr:	Y	His:	H	Gln:	Q
Asn:	N	Lys:	K	Asp:	D	Glu:	E
Cys:	C	Trp:	W	Arg:	R	Gly:	G

10   X: not determined or ambiguous.

We also constructed pBG211-11, a plasmid
coding for the N-terminal 113 amino acids of soluble
T4 protein. This construct, which codes for a pro-
tein characterized by a single disulfide bridge,
15 between the cysteines at amino acid positions 18 and
86, is conveniently expressed in E.coli.

To construct p211-11, as depicted in Fig-
ure 24, we first cut p195-8 (see Figures 8D and 9A)
with ClaI to remove the ClaI-ClaI cassette contain-
20 ing the cDNA sequence of rsT4.2. We then digested
pAT153γ3SH16ΔAmp, the tryptophan operon promoter
plasmid from the gamma interferon producing E.coli
strain BN374 with ClaI, and deleted the cDNA coding
for gamma interferon. Subsequently, we inserted
25 the ClaI-ClaI cassette into the ClaI-cut E.coli
plasmid in front of the tryptophan operon promoter
and ligated to produce p196-10.

As shown in Figure 25, we then subjected
pBG380 to oligonucleotide-directed mutagenesis to
30 insert three tandem translational stop codons follow-
ing the T4 cDNA sequence coding for amino acids -23
to 113 in pBG380, to produce pBG394.

We then constructed p211-11 from fragments
of each of p196-10, pBG394 and p1034 as depicted in
35 Figure 26. The first fragment including the vector
sequences, was produc d by restricting p196-10 with

HindIII and ClaI to remove the T4 coding sequence
from amino acids 61 through 374 of rsT4.2 and includ-
ing vector sequence following the 3' end of the rsT4
gene. The second fragment, a HindIII - BglII segment

5   including the codons for T4 amino acids 61-113 of
rsT4.9 immediately followed by a triplet of stop
codons in tandem, was isolated by HindIII/BglII diges-
tion of pBG394. The third fragment, a BamHI - ClaI
fragment containing a bacteriophage T4 transcriptional

10   termination signal [H. N. Kirsch and B. Allet,
"Nucleotide Sequences Involved In Bacteriophage T4
Gene 32 Translational Self-Regulation", Proc. Natl.
Acad. Sci. USA, 79, pp. 4937-41 (1982)], was isolated
by BamHI/ClaI digestion of p1034. We then ligated

15   these three fragments to produce p211-11, a T4 con-
struct coding for a 113 amino acid soluble form of
T4 protein, with asparagine at amino acid position 3
(i.e., rsT4.113.1).

We then subjected p211-11 to oligonucleo-
20   tide site-directed mutagenesis (Figure 27) to change
the amino acid at position 3 from asparagine to
lysine using the oligonucleotide T4-66:

25
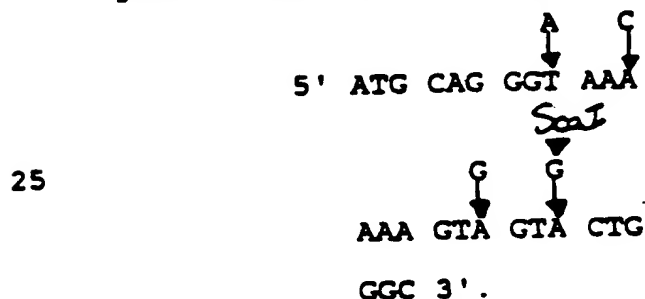
This produced plasmid p214-10, a fully
corrected 113 amino acid soluble T4 vector coding
30   for a 113 amino acid soluble form of T4 protein,
with lysine at amino acid position 3 (i.e.,
rsT4.113.2). As shown in Figure 27, we subjected
p214-10 to oligonucleotide site-directed mutagene-
sis to delete glutamine and glycine at, respectively,

amino acid positions 1 and 2 of the T4 sequence using
the oligonucleotide T4AID-87:

$$\overset{\displaystyle C}{\overset{\displaystyle \downarrow}{}}$$

5' GTA TCG ATT TGG

5   ATG ATG AAA AAA

      GTA GTA 3'.

      This produced p215-7, a 111 amino acid
soluble T4 construct, including the trp promoter,
which directs the expression of a 111 amino acid

10   soluble form of T4 protein, with lysine at amino
acid position 3 (i.e., rsT4.111).

      We next constructed p218-8, a 111 amino
acid construct which directs the expression of a 111
amino acid soluble form of T4 protein, with lysine

15   at amino acid position 3 (i.e., rsT4.111) under the
control of the P_L promoter, as depicted in Figure 28.

      More specifically, we cut p197-12 (Figure
9A) with ClaI to remove the 101 bp fragment contain-
ing linker and terminator sequences. We also cut

20   p215-7 with ClaI to remove the ClaI - ClaI cassette
containing the cDNA sequence of rsT4.111 and the φT4
transcriptional terminator sequence [Kirsch and Allet,
supra]. Subsequently, we inserted the ClaI - ClaI
cassette into the ClaI-cut p197-12 to produce p218-8.

25   In order to express rsT4.113.1, we trans-
formed E.coli A89 with p211-11 by conventional
techniques [Maniatis et al. (1982), supra] to form
E.coli A89/p211-11. E.coli A89 is a tetracycline
sensitive derivative of E.coli SG936. We isolated

30   E.coli A89 from E.coli SG936 according to the method
of S. R. Maloy and W. D. Nunn, "Selection For Loss
Of Tetracycline Resistance By Escherichia coli",
J. Bact., 145, pp. 110-12 (1981), which is based
upon the ability of the lipophilic chelating agent

35   fusaric acid t selectively inhibit resistant strains.

-61-

More specifically, we plated E.coli SG936 on medium
containing, per liter, 5 g tryptone, 5 g yeast extract,
10 g NaCl, 10 g NaH$_2$PO$_4$·H$_2$O, 50 mg chlortetracycline-
HCl, 12 mg fusaric acid, 0.1 mM ZnCl$_2$ and 15 g agar.

5   Colonies which grew at 30°C (putative tetracycline-
sensitive strains) were retested for tetracycline
sensitivity on L-agar plates containing 5 µg/ml
tetracycline. One tetracycline-sensitive strain,
designated A89, was then shown to be unable to grow

10   on LB agar at 42°C, thus verifying the presence of
the htpR mutation.

     Transformants were selected by tetracycline
resistance. We picked a single colony into 20 ml of
minimal medium plus 0.2% casamino acids plus trypto-

15   phan (100 µg/ml) plus tetracycline (10 µg/ml) in a
100 ml shake flask placed in a shaking air incubator
at 30°C and allowed the cells to grow up overnight.
The following morning, we inoculated 40 ml of mini-
mal medium plus 0.2% casamino acids plus tryptophan

20   (100 µg/ml) plus tetracycline (10 µg/ml) with the
overnight culture at OD$_{600}$ = 0.05 in a 500 ml flask.
The cells were grown to midlog phase and then induced
by pelleting, washing once in minimal medium and
then resuspending in minimal medium plus 0.2% cas-

25   amino acids plus tetracycline (10 µg/ml), in the
absence of tryptophan. We removed 0.6 OD$_{600}$ of
cells after 0, 1, 2, 3 and 4 hours incubation and
after growth overnight.

     The aliquots were centrifuged and cell

30   pellets were subjected to lysis by boiling in
Laemmli gel loading buffer. After centrifugation to
remove cell debris, half of each sample was subjected
to SDS-PAGE, followed by Western blot analysis with
our rabbit antip ptid antibody probes or by Coomassie

35   blue protein staining (Figures 29A and 29B).

-62-

Purification Of rsT4.113.1

We then purified rsT4.113.1 from the E.coli
transformant by means of two essentially quantitative
steps involving anion-exchange and gel-filtration
5   chromatographies performed under reducing and dena-
turing conditions.

More specifically, we suspended 14 g of
wet cells from a 4 L shake-flask fermentation in
100 ml of a 20mM Tris (pH 7.5) buffer containing
10   20 µg/ml DNase, 20 µg/ml RNase and 1 mM phenylmethyl-
sulfonylfluoride ("PMSF"). The suspension was applied
to a French Press at 1000 psi in two passages and
then centrifuged in an SA 600 rotor at 18,000 g for
15 min at 4°C. The resulting pellet was solubilized
15   in 20 ml of a 20 mM Tris (pH 7.5) buffer containing
7 M urea and 10 mM 2-mercaptoethanol. We then sub-
jected the suspension to ultracentrifugation at
85,000 g for 90 min at 4°C. The supernatant was
diluted by the addition of 80 ml of 20 mM Tris
20   (pH 7.5) buffer containing 7 M urea and 10 mM
2-mercaptoethanol and 40 ml of the sample was
applied to a 3 x 4 cm Q-Sepharose fast-flow column
(Sigma, St. Louis, Missouri) which had been pre-
equilibrated in the same buffer. The column was
25   developed with a gradient in 400 ml total volume of
increasing NaCl from 0 to 0.3 M in the same Tris/urea/
2-mercaptoethanol buffer. Column fractions were
monitored for absorbance at 280 nm and for protein
content by SDS-PAGE (15% acrylamide). The fractions
30   were also analyzed by Western blots. Figure 30,
panel (a) is a chromatogram displaying the purifi-
cation of rsT4.113.1 by ion-exchange chromatography.
In that figure, peaks containing rsT4.113.1 are
identified. The rsT4.113.1 was found to elute early
35   in the NaCl gradient and to be well-resolved from
low-m lecular weight contaminants.

In order to separate rsT4.113.1 from high-
molecular weight contaminants, we carried out gel-
filtration chromatography on an rsT4.113.1-containing
pool for final purification of the protein to near
homogeneity (>95% purity). More specifically, we
prepared a pool containing 20 mg of protein in 50 ml
and then concentrated to 10 ml in a stirred-cell
ultrafiltration unit (Amicon, Danvers, MA.) using a
PM-30 membrane (Amicon). Subsequently, 5.0 ml of
the concentrate was applied to a 1.5 x 95 cm S-300
column (Sigma) equilibrated and developed in the
same Tris/urea/2-mercaptoethanol buffer. We moni-
tored the column fractions for absorbance at 280 nm
and for protein content by SDS-PAGE. The fractions
were also analyzed by Western blots. A pool con-
taining rsT4.113.1 (approximately 4 mg) in 15 ml was
thus prepared. Figure 30, panel (b) is a chromato-
gram displaying the purification of rsT4.113.1 by
gel-filtration separation of the rsT4.113.1 pool.
In that figure, peaks containing rsT4.113.1 are
identified.

Figure 30, panel (c) is an SDS-PAGE analysis
depicting the purification of the rsT4 derivative
throughout the centrifugation and chromatography
steps. In Figure 30, panel (c), the lanes depicted
are:

lane A: molecular weight standards
lane B: cell extracts
lane C: cell pellet following solubilization
          of cell extract in non-denaturing
          conditions
lane D: supernatant following solubilization
          of cell extract in non-denaturing
          buffer
lane E: supernatant following ultracentri-
          fugation st p

-64-

lane F: Q-Sepharose pool
lane G: S-300 gel-filtration pool.

Refolding Of Purified rsT4.113.1

5   We refolded the purified rsT4.113.1 by
dilution and dialysis steps to non-denaturing and
oxidized conditions. More specifically, refolding
of the protein at a concentration of 0.5 OD (280)/ml
was achieved by stepwise dialysis against 500 volumes
of 3 M urea, 20 mM Tris (pH 7.5); 500 volumes of 1 M
10 urea, 0.1 M ammonium acetate (pH 6.8) and, finally,
the same volume of a phosphate-buffered saline solu-
tion. Throughout the refolding procedure, samples
of the protein were monitored for relative content
by spectral analysis and by high-performance liquid
15 chromatography ("HPLC") performed on a 150A liquid
chromatographic system (Applied Biosystems, Inc.,
Foster City, California). An octasilyl column
(Aquapore RP-300, 0.46 x 3.0 cm) was equilibrated in
80% 0.1% trifluoroacetic acid ("TFA")/water (sol-
20 vent A) and 20% 0.085% TFA/70% acetonitrile (sol-
vent B) and developed with a linear gradient of
increasing acetonitrile concentration from 20% to
80% (solvent B) over 45 min at a flow rate of
0.5 ml/min.

25   As shown in Figure 31, panel (a), protein
in 7 M urea, 10 mM 2-mercaptoethanol and 20 mM
Tris(pH 7.5) eluted from the HPLC column at 49%
acetonitrile in the gradient. In subsequent steps,
from 1 M urea/1 mM ammonium acetate (pH 6.8) [Fig-
30 ure 31, panel (b)] to phosphate buffered saline
[Figure 31, panel (c)], an increasing percentage of
rsT4.113.1 was found to elute earlier in the HPLC
gradient -- at 47% acetonitrile. The identity of
the earlier eluting peak as oxidized product was
35 verified by reduction f rsT4.113.1 in non-chaotropic

solutions and application of sample thus treated to
HPLC under the same conditions.

The elution of oxidized rsT4.113.1 prior to
reduced protein on HPLC suggests that formation of
5   the single disulfide bridge decreases relative hydro-
phobicity of the protein [J. L. Browing et al., _Anal._
Biochem., 155, pp. 123-28 (1986)]. Spectral analysis
of rsT4.113.1 was performed throughout the course of
refolding in order to monitor relative yield of solu-
10 ble protein in the procedure. The refolding method
allowed approximately 20% recovery of rsT4.113.1.
HPLC analysis indicated a less than 15% contaminant
of reduced protein in the preparation (Figure 30,
panel (c), lane G).

15 <u>Sequencing Of Renatured rsT4.113</u>

We then carried out amino acid analysis of
rsT4.113.1 by automated Edman degradation in an
Applied Biosystems 470A gas phase sequenator equipped
with a 900 A data system. Phenylthiohydantion amino
20 acids generated during the course of the degradative
chemistry were analyzed on-line using an Applied
Biosystems 120A PTH-analyzer equipped with a PTH-C18
2.1 x 220 mm column. Protein (10 µg) for sequence
analysis was applied to SDS-PAGE (15% acrylamide)
25 and electroblotted on an Immobilon membrane (Millipore
Corp., Bedford. Massachusetts) as described by
P. Matsudaira, _J. Biol. Chem._, 262, pp. 10035-38
(1987).

Amino acid analysis of protein samples was
30 performed by hydrolysis of protein in 6 N HCl, in
vacuo, for 24 h at 110°C. The hydrolysates were
then applied to a Beckman 6300 Analyzer equipped
with post-column detection by ninhydrin. Western
blot analysis of the SDS-PAGE gels was carried out
35 by standard techniques using rabbit antisera JB-1.

-66-

Sequence analysis revealed an amino terminal sequence
of: Met-Gln-Gly-Asn-Lys-Val-Val ...

The purified rsT4.113.1 protein was found
to contain stoichiometric quantities of the amino-
5   terminal methionine placed in the protein construct
for expression in E.coli and an intact polypeptide
chain consistent with a sequence derived from the
plasmid construction. Recovery of phenylthiohydan-
toinyl-methionine at the first cycle of the degrada-
10   tive chemistry was 60% consistent with routine initial
yields obtained in the automated Edman. This obser-
vation excludes the possibiity that a significant
percentage of the rsT4.113.1 lacked the initiation
methionine, i.e., the NH_2-methionine was not removed
15   by expression of rsT4.113.1 in E.coli, or that sequence
analysis was impaired by the presence of glutamine
at the first cycle of the degradative chemistry.
Sequence analysis was performed for 40 cycles and no
evidence of lysine carbamylation was observed. Amino
20   acid analysis displayed a close correlation of actual
and theoretical values for amino acids, thus indicating
the marked absence of proteolytic degradation in the
course of expression, or purification, or both.

Immunoprecipitation Of CHO Cell
25   Lines Producing Soluble T4

We tested the conditioned media from ^{35}S-Cys
metabolically labelled CHO cells transfected with
one of the T4 mutant constructs pBG377, pBG380, pBG381,
the full length recombinant T4 construct pBG379, of
30   this invention or vector only, to determine whether
any produced a molecule recognized by the anti-T4
monoclonal antibody 19 Thy. To carry out this test,
we incubated about 10^7 CHO cells transfected with
either pBG380, pBG381, pBG377, pBG379 or pBG312, for
35   5 hours at 37°C with 180 μCi/ml ^{35}S-labelled cysteine

-67-

[DuPont, New England Nuclear] in 4 ml RPMI cys⁻ medium
(Gibco). After labelling of the cells, 1 ml of fil-
tered, conditioned media was made 0.5 mM with phenyl-
methyl-sulphonyl fluoride and immunoprecipitated

5   with OKT4 and protein A Sepharose [P. E. Sayre and
E. L. Reinherz, *Eur. J. Immunol.*, 15, pp. 291-95
(1985)]. Subsequently, we incubated media from the
^{35}S-labelled cells with OKT4 (ATCC #CRL 8002). We
then immuno-precipitated with protein A Sepharose

10   and subjected the immuno-precipitates to SDS-PAGE
under reducing conditions on 10% polyacrylamide gels
[U. K. Laemmli, *Nature*, 227, pp. 680-85 (1980)].
Autoradiography was carried out with X-Omat X-ray
film (Eastman Kodak).

15   As shown in lanes 3-5 of Figure 32, both
pBG380 (rsT4.2) and pBG381 (rsT4.3) directed the
synthesis of a secreted, immune, ^{35}S-labelled T4
protein that was recognized by the OKT4 anti-T4
antibody. The immunoprecipitated truncated mole-

20   cules migrated as 49 Kd proteins, a result consis-
tent with their predicted molecular weights. In
contrast, no soluble T4 antigen could be detected in
the conditioned media of cell lines stably trans-
fected with pBG377 (rsT4.1) or pBG379 (rflT4).

25   Immunoprecipitation analysis of cellular extracts of
cell lines transfected with pBG377 suggests that the
rsT4.1 gene may be misfolded, which could account
for a block in its secretion [M. J. Gething et al.,
Cell, 46, pp. 939-50 (1986)].

30   In Figure 32, the lanes represent the
following: Lane 1: immunoprecipitation from condi-
tioned medium of CHO cells stably co-transfected
with vectors pBG312 and pAdD26. Lane 2: blank.
Lanes 3 and 4: immunoprecipitation from conditioned

35   medium of CHO cells stably co-transfected with pBG380
(rsT4.2) and pAdD26. Lanes 5 and 6: immunopr cipita-
tion from conditioned medium of CHO cells stably

co-transfected with pBG381 (rsT4.3) and pAdD26.
Lane 7: immunoprecipitation from conditioned medium
of CHO cells stably co-transfected with recombinant
full length T4 (pBG379) and pAdD26. In Figure 32,
5   the arrow indicates the predicted position of the
soluble T4 from pBG380 or pBG381 relative to the
migration of standard molecular weight markers.

Immunoprecipitation Of COS 7 Cell Lines
Producing Recombinant Soluble T4

10   We expressed recombinant soluble T4
derivatives pBG392, pBG393 and pBG394 in COS 7 cells
by electroporation, essentially as described by
G. Chu et al., "Electroporation For The Efficient
Transfection Of Mammalian Cells With DNA", Nuc.
15   Acids Res., 15, pp. 1311-26 (1987). More specifi-
cally, we introduced 20 μg closed circular plasmid
DNA and 380 μg of carrier (sonicated salmon sperm
DNA) into 3×10^7 COS 7 cells. The cells were
electroporated using a Gene Pulser (Biorad) set at
20   300 volts. Subsequently, we incubated the COS 7
cells in Dulbecco's Modified Eagle's Medium supple-
mented with 10% fetal calf serum for 24 hours. We
then harvested the conditioned media, filtered it
through a Millipore Millidisk 0.22 μ hydrophilic
25   filter cartridge (Millipore #MCGL 305-01) and
concentrated the secreted proteins on a fast-S ion
exchange column (S-Sepharose Fast Flow, Pharmacia
#17-0511-01) in 20 mM MES buffer (pH 5.5).
          We then eluted the bound proteins with
30   20 mM Tris-HCl (pH 7.7) and 0.3 M NaCl. The elution
pool was subsequently diluted with 2 volumes of 20 mM
Tris-HCl (pH 7.7) and it was then loaded on a column
comprising either 19Thy anti-T4 monoclonal antibody
and protein A Sepharose or OKT4A and protein A
35   Sepharose. We washed the column extensively and
eluted the bound material as 0.5 ml fractions with

50 mM glycine-HCl (pH 2.5), 150 mM NaCl, 0.1 mM EGTA
and 5 µg/ml Bovine pancreatic trypsin inhibitor,
Aprotinin (Sigma, #A1153). The immunoprecipitates
were subjected to SDS PAGE (10% gel) followed by
5   immunoblotting against rabbit antisera raised
against peptide JB-1. We employed silver stained
gels to follow binding and elution of rsT4 during
chromatography.

Figure 33 depicts an immunoblot analysis of
10   transiently expressed pBG392 (rsT4.7) [lanes 10,
11]; pBG393 (rsT4.8) [lanes 4, 7, 8] and pBG394
(rsT4.9) [lane 5]. The standards are 50 ng purified
rsT4.3 (lane 1); 150 ng purified rsT4.3 (lane 2) and
250 ng purified rsT4.3 (lane 3). The arrow indicates
15   the expected position of migration of a protein with
the relative molecular weight of rsT4.7: 21,000
daltons. The sample that was to be loaded into lane 4
was lost and lanes 6 and 9 are blank.

As shown in lanes 10 and 11 of Figure 35,
20   pBG392 (rsT4.7) directed the synthesis of a secreted,
immune protein that was recognized by the anti-T4
antibodies OKT4A and 19Thy. Lanes 4, 7 and 8 also
demonstrate that pBG393 (rsT4.8) directed the
synthesis of a secreted, immune protein that was
25   recognized by OKT4A and 19Thy. This analysis
illustrates that rsT4.7 contains the OKT4A epitope.
It also suggests that the binding region for HIV
envelope binding resides in the amino 182 terminal
residues of T4.

In contrast, no soluble T4 could be detected
30   in the media of cell lines transfected with pBG394
(rsT4.9) [see lane 5]. Immunoprecipitation analysis
of cellular extracts of cell lines transfected with
pBG397, however, showed that rsT4.9 was recognized
by OKT4A. We believe that rsT4.9, a 113 amino acid
35   construct, binds the HIV virus and that it represents
a second generation soluble T4, one with only two

cysteines and one of three disulfide bridges. Accordingly, rsT4.9 is easily produced in E.coli or yeast systems.

Similarly, although no soluble T4 could be
5   detected in the media of cell lines transfected with pBG396 (rsT4.12), analysis of cellular extracts of those cell lines showed that rsT4.12 was recognized by OKT4A. Thus, rsT4.12 may also bind HIV virus.

Radioimmunoassay And Epitope
10 Analysis Of rsT4.113

In order to determine if the 113 fragment of rsT4 contained structural determinants for binding to OKT4A, Leu-3A and OKT4, we then carried out radioimmunoassay and epitope analysis of rsT4.113
15 using a competitive inhibition radioimmunoassay [C. J. Newby et al., "Solid-Phase Radioimmune Assays" in Handbook Of Experimental Immunology, D. M. Weir (Ed.), 1, pp. 34.1-34.8 (1986)]. As OKT4A and Leu-3A block infectivity of HIV in vitro [Dalgleish et al.,
20 supra] and binding of T4 to gp120/160 [McDougal et al., supra], this analysis served as a first approximation as to whether or not rsT4.113 contained structural elements for interaction with HIV.

We first coated U-bottom 96 well microtiter
25 plates (Falcon) with 50 µl/well goat-anti-mouse IgG (Hyclone Typing Kit, Logan, Utah) in PBS (pH 7.0) to a concentration of 50 µg/ml and incubated the plates overnight at 4°C. We then rinsed the plates with 1X PBS and blotted them dry. The plates were then
30 blocked by the addition of 100 µl/well of a 1X PBS solution containing 5% bovine serum albumin for 1 hour at room temperature. We rinsed the plates with PBS, blotted dry and then spotted them with 50 µl of one of three antibody solutions containing
35 either OKT4 (10 µg/ml in block buffer); OKT4A (500 ng/ml in block buffer) r Leu-3A (B ct n-

-71-

Dickinson) (500 ng/ml in block buffer). We let the plates stand for 2 hours at room temperature. We then washed the plates 3 times with a PBS/0.05% Tween-80 solution and 2 times with 1X PBS and blotted them dry.

5

In a separate plate, we titrated competitor samples of unlabeled rsT4.113.1 from 20 μg/ml and serially diluted twice (including no competitor control), with final volumes in each well of 25 μl.

10   The positive control for this assay was competition with unlabeled rsT4.3 (375 amino acids). We then added 25 μl of ^{125}I-rsT4.3 containing 10,000 cpm/25 μl (prepared according to A. E. Bolton and W. M. Hunter, Radioimmunoassay And Related Methods,

15   Chapter 2c). Subsequently, we spotted the entire 50 μl content of each well onto the assay plate containing each of the three antibody solutions and incubated for 2 h at room temperature. We then washed the plates 3 times with a PBS/0.5% Tween-80

20   solution and 2 times with 1X PBS, blotted them dry and then counted the wells in a Beckman gamma counter for radioactivity.

As shown in Figure 34, rsT4.113.1 competes with ^{125}I-rsT4.3 for absorption to an OKT4A solid phase in a dose-dependent manner. Additionally,

25   rsT4.113.1 competes with ^{125}I-rsT4.3 for absorption to a Leu-3A solid phase in a dose-dependent manner. By comparison to unlabeled rsT4.3, rsT4.113.1 exhibits a molar affinity for those antibodies within a factor

30   of 3. In the 0.4 to 25 μg/ml concentration range tested, rsT4.113 did not compete with radiolabelled rsT4.3 for binding to OKT4. In a similar assay, we observed that rsT4.111 also competes with ^{125}I-rsT4.3 for binding to OKT4A and Leu-3A, but not to OKT4

35   [Figures 35-37].

Based on thes results, w b lieve that the epitopes for OKT4A and Leu-3A are contained within

the amino-terminal 113 amino acids of T4. We also believe that the epitope for OKT4 binding is localized within the carboxy terminal of the T4 polypeptide.

Accordingly, we believe that the gp120-binding domain is localized within the amino terminal 113 or 111 amino acids of the T4 protein. Based on this belief, we synthesized various synthetic oligo-peptides which contain sequence within that structural domain. These oligopeptides are represented in Figure 3 as follows:

Oligopeptide	Amino Acid Coordinates
JB-1	44-63
rsT4 #6	18-29
rsT4 #7	5-56
rsT4 #8	84-97
rsT4 #9	30-63

We synthesized these peptides using conventional phosphoamide DNA synthesis techniques [Tetrahedron Letters, 22, pp. 1859-62 (1981)]. We synthesized the peptides on an Applied Biosystems 380A DNA Synthesizer and purified them by gel electrophoresis.

ELISA Assay For rsT4.113

We also carried out an ELISA assay for rsT4.113.1 produced by p211-11-transformed E.coli. Throughout this assay, dilutions were made in block-ing solution and, between each step, we washed the plates with PBS/0.05% Tween-20. More specifically, we coated wells of Immulon 2 (Dynatech, Chantilly, Virginia) plates with .005 OD (280 nm)/ml of OKT4 (IgG2b) in 0.05 M bicarbonate buffer to a volume of 50 µl/well and incubated the plates overnight at 4°C. We then blocked the plates with 5% bovine serum albumin in PBS, 200 µl/well, and incubated for 30 minutes at room temperature.

Subsequently, we added 50 µl of 50 ng/ml rsT4.3 to each w 11, incubating overnight at 4°C.

-73-

We then added 50 μl/well of a mixture containing
rsT4.113.1 and 10 ng/ml of OKT4A and incubated for
2 1/2 hours at room temperature. Using a Hyclone
Kit (Hyclone), we then carried out the following

5 steps. First, we added 1 drop of rabbit anti-mouse
IgG2a to each well and incubated the plates for
1 hour at room temperature. We then added 100 μl of
peroxidase-labeled anti-rabbit IgG, diluted 1:4000
with blocking buffer to each well, and incubated for

10 1 hour at room temperature.

We prepared a substrate reagent as follows.
We diluted substrate reagent 1:10 in distilled water
and added two O-phenyl-ethylene-diamine ("OPD")
chromophore tablets per 10 ml of substrate. We let

15 the mixture dissolve thoroughly by mixing with a
vortex. Alternatively, a TMB peroxidase substrate
system (Kirkegaard & Perry Catalogue #50-76-00) may
be used. Subsequently, we added 100 μl of the
chromophore solution to each well, incubated for

20 10-15 minutes at room temperature and then stopped
the color development with 100 μl of 1N H_2SO_4. We
then measured OD at 490 nm, using an ELISA plate
reader.

The results of the assay are demonstrated

25 in Figure 38.

We then subjected the soluble T4 proteins
produced by the T4 constructs of this invention to
various functional assays.

<u>Assays Of The Antiviral Activity Of Soluble T4</u>

30   The antiviral activity of soluble T4
according to this invention was evaluated using
modifications of various <u>in vitro</u> systems used to
study antiviral agents and neutralizing antibodies
[D. D. H et al., "Recombinant Human Interferon Alpha

35 (A) Suppresses HTLV-III Replication <u>In Vitr</u>",
<u>Lancet</u>, pp. 602-04 (1985); K. Hartshorn et al.,

-74-

"Synergistic Inhibition Of HTLV-III Replication
In Vitro By Phosphonoformate And Recombinant Inter-
feron Alpha-A", Antimicrob Ag Chemoth, 30, pp. 189-91
(1986)].

5   For each of these assays, we prepared graded
concentrations of soluble T4 and preincubated them
with an H9 derived IIIB isolate of HIV [a gift from
Drs. M. Popovic and R. Gallo, National Cancer
Institute, Bethesda, Maryland]. The isolate was
10 maintained as a chronically infected culture in H9
cells. Cell-free HIV stocks were obtained from
supernatant fluids of HTLV-III infected H9 cultures
(culture conditions: 1×10^6 cells/ml with 75% viable
cells). We prepared serial 10 fold dilutions of
15 recombinant soluble T4 ranging from 10 picograms/ml
to 10 micrograms/ml and incubated them with fifty
50% tissue culture infectious doses ($TCID_{50}$) of HIV
for 1 hour at 37°C, in RPMI-1640 supplemented with
20% heat inactivated fetal calf serum (FCS). We
20 then added 150 µl of H9 cells to a final concentra-
tion of 0.5×10^6 cells/ml which were not HIV-infect d
to the wells containing aliquots of the recombinant
soluble T4/HIV mixture.

           We adjusted each virus inoculum to a con-
25 centration of 250 $TCID_{50}$/ml. We preincubated 100 µl
of the virus inoculum with 200 µl recombinant solu-
ble T4 or 100 µl immunoglobulin prepared in tripli-
cate serial 2-fold dilutions for 1 hour at 37°C
prior to inoculation onto $1.5 - 2 \times 10^6$ H9 cells in
30 5 ml RPMI 1640 supplemented fetal calf serum (20%),
HEPES (10mM), penicillin (250 U/ml), streptomycin
(250 µg/ml) and L-glutamine (2mM). On days 5, 6, 7,
10 and 14, we examined each culture for characteris-
ic cytopathic effects ("CPE"). Neutralization was
35 defined as the inhibition of syncytia formation com-
ared with c ntr ls.

-75-

The positive control used was HIV seropositive neutralizing serum, as described in D. D. Ho et al., "Human Immunodeficiency Virus Neutralizing Antibodies Recognize Several Conserved Domains On
5 The Envelope Glycoproteins", J. Virol., 61, pp. 2024-28 (1987). The negative controls used were HIV seronegative serum only and buffer only.

Cytopathic Effect Assay (CPE)

In this assay, following conventional
10 protocols for cytopathic effect assays [Klatzmann et al. (1984), supra and Wong-Staal and Gallo (1985), supra], we microscopically examined the H9 cells for evidence of cytopathic effects of HIV.
The CPE was scored on a four point scale
15 from 1+ to 4+, with 4+ representing the highest degree of CPE.
By day 14, wells containing recombinant soluble T4 according to this invention (rsT4.2, derived from the pBG380 transfected CHO cell line
20 BG380) at 10 µg/ml showed no evidence of CPE, while the negative control showed 1+ to 3+ CPE.

p24 Radioimmunoassay

We then tested soluble T4 as an inhibitor of viral replication in an HIV virus replication
25 assay according to D. D. Ho et al., J. Virol., 61, pp. 2024-28 (1987) and J. Sodroski et al., Nature, 322, pp. 470-74 (1986). We carried out the assay essentially as described, except that the cultures were propagated in microtiter wells containing
30 200 µl. In this assay, we evaluated the ability of the soluble T4 polypeptides of this invention to block HIV replication, as measured by HIV p24 antigen production. We sampled supernatants twic w ekly for HIV p24 antigen as described below.

We obtained an assay kit [HTLV-III p24 Radioimmunoassay System, Catalogue No. NEK-040, NEK-040A, Biotechnology Systems, New Research Products, Dupont] which contains affinity purified

5   ^{125}I labelled HIV p24 antigen, a rabbit anti-p24 antibody and a second goat anti-rabbit antibody which is used to precipitate antigen-antibody complexes. We carried out the assay according to the protocol included with the kit. Accordingly, we mixed a sample

10   to be assayed or one of a series of amounts of unlabelled p24 antigen with a fixed amount of ^{125}I labelled p24 and a fixed limited amount of rabbit anti-p24 antibody. We incubated the samples over-night at room temperature and then added a goat

15   anti-rabbit immunoglobulin preparation for 5 minutes at 40°C. We centrifuged the samples in a microfuge and aspirated the supernatant fluid. Pelletted ^{125}I labelled p24 was quantitated for each sample by gamma counting and a standard curve for the ^{125}I p24 dis-

20   placed by the known amounts of antigen added to standard tubes was constructed. We then calculated the ^{125}I labelled p24 displaced by the antigen present in the unknown samples by interpolation using the standard curve constructed from the known amounts of

25   p24 antigen contained in the standard samples. The results are shown in the table below.

-77-

p24 ASSAY OF HIV REPLICATION INHIBITION

Day	rsT4.2 (µg/ml)	Patient Serum	Average CPM	% Bound/ Unbound
7	-	Negative	344	8.5
	-	Positive	2,237	112.4
	0.5*	-	551	19.9
	5.0**	-	1,766	86.6
10	-	Negative	230	2.2
	-	Positive	2,459	124.6
	0.5*	-	322	7.3
	5.0**	-	1,980	96.3
14	-	Negative	221	1.8
	-	Positive	2,284	115.0
	0.5*	-	246	3.1
	5.0**	-	1,988	98.7

These results demonstrate that soluble T4
according to this invention at a concentration of
5 µg/ml completely inhibits virus replication as
measured in this standard 14 day assay. These
results are also depicted in Figure 39 in graphic
form. In Figure 39, values were calculated from a
standard curve of p24 according to assay kit
instructions.

---

* This concentration was initially believed to be
1.0 µg/ml, based upon our preliminary approximation
that 1 unit of absorbance at 280 nm ("A_{280}"), was
equivalent to 1 mg of rsT4.2. Absorbance at 280 nm
is a commonly used first approximation of protein
concentration. Upon amino acid analysis of the pro-
tein, however, we found that it had a higher extinc-
tion coefficient than originally approximated, with
1 A_{280} unit of rsT4.2 being equivalent to 0.5 mg of
the protein.

** This concentration was initially believed to
be 10 µg/ml, based upon our preliminary approximation
that 1 unit of absorbance at 280 nm ("A_{280}"), was
equivalent to 1 mg of rsT4.2. Absorbance at 280 nm
is a commonly used first approximation of protein
concentration. Upon amino acid analysis of the pro-
tein, however, w found that it had a higher extinc-
tion coefficient than originally approximated, with
1 A_{280} unit of rsT4.2 being equival nt to 0.5 mg of
the protein.

We then carried out a p24 replication assay
as described above, except that the soluble T4 was
added to the infected cultures during refeeding at
days 3, 7 and 10, in order to maintain a constant
5   rsT4 concentration throughout the infection period.
The results of this assay are shown in the table
below.

INHIBITION OF HIV REPLICATION WITH CONSTANT CONCENTRATION OF rsT4

rsT4.2 (µg/ml)	p24 (ng/ml)
0.008	770
0.031	970
0.125	85
0.5	0
5.0	0
0	1120
uninfected	0

These results demonstrate that when solu-
20   ble T4 protein according to this invention was main-
tained at a constant concentration throughout the
infection period, as little as 0.125 µg/ml of the
protein substantially blocked replication of 250
$TCID_{50}$/ml of HIV-1.
25   Advantageously, soluble T4 protein accord-
ing to this invention, at concentrations far exceed-
ing those required to block viral replication, did
not exert immunotoxic effects _in vitro_, as measured
by three lymphocyte proliferation assays -- mixed
30   lymphocyte response, phytohemagglutinin, and tetanus
toxoid stimulated response.

Syncytia Inhibition Assay

To further assess the effect of soluble T4
on HIV env-T4 binding, we evaluated the effect of two
35   preparations of our soluble T4 protein on the syn-
cytiagenic properties of HIV in th co-cultivation
assay. We carried out a C8166 cell fusi n assay

-79-

as described in B. D. Walker et al., Proc. Natl.
Acad. Sci. USA, 84, pp. 8120-24 (1984).

We incubated 1×10^9 H9 cells chronically
infected with HTLV-IIIB for 1 hour at 37°C in 5%
5   CO_2 with various concentrations of one of two
preparations of rsT4.2 in 150 µl RPMI-1640 media
supplemented with 20% fetal calf serum. We then
added 3×10^4 C8166 cells in 50 µl media (a T4$^+$
transformed human umbilical cord blood lymphocyte
10   line [Sodroski et al., supra], to a final volume of
0.2 ml in each well. Final well concentrations of
soluble T4 were 0.5 µg/ml* and 5.0 µg/ml* for prepa-
ration #1 and 1.25 µg/ml* and 12.5 µg/ml* for prepa-
ration #2. We then counted total number of syncytia
15   per well at 2 hours and 4 hours after adding the
C8166 cells at 37°C in 5% CO_2. Parallel co-cultiva-
tions used buffer alone (negative control) or OKT4A
at 25 µg/ml (positive control) as controls. We con-
sidered a positive result as a 50% reduction in
20   syncytia compared to controls, at a time when at
least 100 syncytia per 10^4 infected H9 cells were
present in the control cultivations. The results of
this assay are shown below and in Figure 40 (2 hour
data).

25   _____

*   These concentrations were initially believed
to be, respectively, 1 µg/ml, 10 µg/ml, 2.5 µg/ml
and 25 µg/ml, based upon our preliminary approxima-
tion that 1 unit of absorbance at 280 nm ("A_{280}"),
30   was equivalent to 1 mg of rsT4.2. Upon amino acid
analysis of the protein, however, w found that it
had a higher extinction co fficient than originally
approximated, with 1 A_{280} unit of rsT4.2 being
equivalent to 0.5 mg of the protein.

-80-

INHIBITION IN C8166 FUSION ASSAY

Preparation	[rsT4.2] (µg/ml)	% Inhibition* 2 Hrs	% Inhibition* 4 Hrs
buffer	0	0	0
rsT4.2	0.5**	30	42
rsT4.2	5.0**	54	47
rsT4.2	1.25**	16	21
rsT4.2	12.5**	77	55
OKT4A (25 µg/ml)	0	100	100

As demonstrated in this table and in Fig-
ure 40, soluble T4 according to this invention at
5.0 µg/ml and 12.5 µg/ml inhibited syncytia formation
at 2 hours, as compared to buffer alone. By 4 hours
after the addition of C8166 cells, soluble T4 at
12.5 µg/ml continued to inhibit greater than 50%
syncytia formation, as compared to the negative
control.

We also evaluated the effect of two prep-
arations of our soluble T4 protein rsT4.7 on the
syncytiagenic properties of HIV in a similar co-
cultivation assay. The results of this assay are
shown below.

---

*   All assays were carried out in triplicate, and
the number of syncytia counted per well was averaged
to calculate % inhibition. The % inhibition repre-
sents the difference between the average number of
syncytia in the negative control (without rsT4 or
OKT4A) and the average number of syncytia counted
when either rsT4 or OKT4A were present during the
assay, divided by the average syncytia count for
the negative control and multiplied by 100.

**   These concentrations were initially believed
to be, respectively, 1 µg/ml, 10 µg/ml, 2.5 µg/ml
and 25 µg/ml, based upon our preliminary approxima-
tion that 1 unit of absorbance at 280 nm ("A_{280}"),
was equivalent to 1 mg of rsT4.2. Upon amino acid
analysis f the protein however, w found that it
had a higher extinction coefficient than originally
approximated, with 1 A_{280} unit f rsT4.2 being
equivalent t 0.5 mg of the protein.

-81-

INHIBITION IN C8166 FUSION ASSAY

Assay date: day 1

5	Preparation	rsT4.7 (µg/ml)	Average Syncytia/50µl aliquot	% Inhibition at 2 Hrs
	H9 cells (control)	0	0	N/A
	C8166 cells (control)	0	0	N/A
10	HIV-infected H9 cells added to C8166 cells (control)	0	118	0
15	OKT4A (control)	0	0	100
	Prep. 1 of rsT4.7	≅ 5.0*	43	63.6

20   *   This concentration was initially believed to be 10 µg/ml, based upon our preliminary approximation that 1 unit of absorbance at 280 nm ("A_{280}"), was equivalent to 1 mg of rsT4.2. Upon amino acid analysis of the protein, however, we found that it
25   had a higher extinction coefficient than originally approximated, with 1 A_{280} unit of rsT4.2 being equivalent to 0.5 mg of the protein.

-82-

Assay date: day 13

	Preparation	rsT4.7 (µg/ml)	Average Syncytia/50µl aliquot	% Inhibition at 2 Hrs
5	H9 cells (control)	0	0	N/A
	C8166 cells (control)	0	1	N/A
10	HIV-infected H9 cells added to C8166 cells (control)	0	141	0
	OKT4A (control)	0	0	100
15	Prep. 2 of rsT4.7	≅ 5.0*	27	80.9

---

*   This concentration was initially believed to
be 10 µg/ml, based upon our preliminary approxima-
tion that 1 unit of absorbance at 280 nm ("A_{280}"),
20   was equivalent to 1 mg of rsT4.2. Upon amino acid
analysis of the protein, however, we found that it
had a higher extinction c efficient than originally
approximated, with 1 A_{280} un t f rsT4.2 being
  quivalent t 0.5 mg f th rotein.

-83-

Assay date: day 14

	Preparation	rsT4.7 (µg/ml)	Average Syncytia/50µl aliquot	% Inhibition at 2 Hrs
5	H9 cells (control)	0	0	N/A
	C8166 cells (control)	0	0	N/A
10	HIV-infected H9 cells added C8166 cells (control)	0	128	0
	OKT4A (control)	0	0	100
15	Prep. 1 of rsT4.7	≊ 5.0*	35	72.7
	Prep. 2 of rsT4.7	≅ 5.0*	2	98.4

        As demonstrated in these tables, soluble
T4 protein rsT4.7 inhibited syncytia formation in
20 HIV-infected H9 cells.
        We also evaluated the effect of rsT4.113.1
and rsT4.111 on the syncytiagenic properties of HIV in
a co-cultivation assay. We carried out a C8166 cell
fusion assay as described in Walker et al., supra.
25   We incubated 1 x 10^4 H9 cells chronically
infected with HTLV-IIIB for 1 hour at 37°C in 5%
CO_2, with from 5 to 50 µg/ml rsT4.113.1 or rsT4.111
in 150 µl RPMI-1640 media supplemented with 20%
fetal calf serum in 96-well microtiter plates. We

30 _____

*   This concentration was initially believed to
be 10 µg/ml, based upon our preliminary approxima-
tion that 1 unit of absorbance at 280 nm ("A_{280}"),
was equivalent to 1 mg of rsT4.2. Upon amino acid
35 analysis of the protein, however, we found that it
had a higher extinction coefficient than originally
approximated, with 1 A_{280} unit of rsT4.2 being
equivalent t 0.5 mg of the protein.

then added 3 x 10^4 C8166 cells to the wells in 50 µl
aliquots. The plates were incubated for 2 hours at
37°C in 5% CO$_2$ and, following this incubation, the
number of syncytia per well were counted.

Syncytia were defined as cells containing
a ballooning cytoplasm greater than three cell
diameters. All samples were counted twice. Parallel
co-cultivation used OKT4A alone or rsT4.3 alone at a
concentration of 25 µg/ml (positive controls) or H9
cells alone or C8166 cells alone (negative controls).
The results of this assay are shown below and in
Figure 41.

INHIBITION IN C8166 FUSION ASSAY

Preparation	rsT4(µg/ml)	% Inhibition
H9 cells (control)	0	0
C8166 cells (control)	0	0
rsT4.113.1	1.25	35
rsT4.113.1	2.5	63
rsT4.113.1	4.25	63
rsT4.113.1	6.25	82
rsT4.113.1	12.5	96
rsT4.3	12.5	100
OKT4A (25 µg/ml)	0	100

As demonstrated in this table and in
Figure 41, rsT4.113.1 exhibited a dose-dependent
inhibition of HIV-induced syncytia formation. The
molar specific inhibitory activity of rsT4.113.1
appeared to be reduced by an order of magnitude by
comparison to anti-viral activity of longer forms of
recombinant soluble T4. Thus, whereas rsT4.113.1 is
effective toward neutralization of HIV-d pendent
c ll fusion _in vitro_, its molar specific inhibitory

-85-

activity is decreased by a factor of 10. It is
undetermined whether this decreased potency is due
to incomplete renaturation of the E.coli-derived
protein, the presence of three additional amino
acids at the N-terminus of rsT4.113.1 (Met-Gln-Gly)
lacking in rsT4.2 or rsT4.3 produced in mammalian
cells, or the absence of additional structure in
rsT4.113.1 required for high-affinity binding to
HIV.

We also carried out a C8166 cell fusion
assay with rsT4.111, as described for rsT4.113.1.
The results of this assay are shown below.

INHIBITION IN C8166 FUSION ASSAY

Preparation	rsT4(µg/ml)	% Inhibition
H9 cell (control)	0	0
C8166 cells (control)	0	0
rsT4.111	1.25	0
rsT4.111	2.5	40
rsT4.111	4.25	20
rsT4.111	6.25	67
rsT4.111	12.5	100
rsT4.111	25.0	100
rsT4.3	12.5	100
rsT4.3	25.0	100
OKT4A (25 µg/ml)	0	100

As demonstrated in this table, rsT4.111
exhibited a dose-dependent inhibition of HIV-induced
syncytia formation. At a concentration of 12.5 µg/ml
and 25.0 µg/ml, complete inhibition of cell fusion
was achieved.

Kinetics Of Intramuscular Injection Of Soluble T4

We examined the kinetics of the appearance
of a recombinant soluble T4 protein according to
this invention (specifically, rsT4.3 from the pBG381-
transfected cell line BG381) in serum after intra-
muscular inj ction as follows.

We obtained two cynomolgus monkeys (<u>Macaca fascicularis</u>) who were free of infectious disease and in good health. Each monkey had been subjected to a 6 week quarantine period prior to administration
5   of the soluble T4 protein. Throughout the administration period, each monkey was maintained on a conventional diet of monkey chow supplemented with fresh fruit. A catheter and a vascular access port were surgically placed in a femoral vein of each animal
10 prior to treatment in order to facilitate blood collection.

Over a period of 28 days, each animal received recombinant soluble T4 protein twice daily by intramuscular injection to the large muscles of
15 the thighs or buttocks. Injections were administered to each animal 8 hours apart and each injection contained a volume of 0.15 ml/kg (0.25 mg/kg) of rsT4.3 (from the pBG381-transformed cell line BG381), for a total dose of 0.5 mg/kg/day/monkey. Serum samples
20 for clearance determination were collected on day 1 before the first treatment and at 1, 2, 4 and 8 hours after the first injection, as well as 1, 2, 4, 14 and 16 hours after the second injection on days 7, 14 and 28.

25 We found that intramuscularly injected soluble T4 reached the maximum level in serum between 1 and 2 hours after injection, with the level falling off slowly and reaching half-maximum value at approximately 6 hours post-injection. According to data
30 obtained for intravenous administration (not shown), the level of rsT4.3 in serum should drop below that attained via intramuscular injection aproximately 2 hours after intravenous injection. Thus, while the maximum rsT4.3 level in serum after intramuscular
35 injection does not reach that attainable via intravenous injection, it is slowly released int the blood stream, remaining detectable in serum for a

-87-

much longer time. This slow release mechanism asso-
ciated with intramuscular routes of injection is
advantageous because a higher level of soluble T4
protein is available over a longer period of time
over a given concentration; thus remaining in a sus-
tained level. Intramuscular administration of solu-
ble T4 protein is particularly useful in treating
early stage HIV-infected patients, to prevent the
virus from disseminating, or in treating patients
who have been exposed to the virus and who are not
yet seropositive.

We determined serum levels of rsT4.3 using
an ELISA assay. Throughout this assay, dilutions
were made in blocking solution and, between each
step, we washed the plates with PBS/0.05% Tween-20.
More specifically, we coated wells of Immulon 2
plates with .01 OD (280 nm)/ml of OKT4 (IgG2b) in
0.05 M bicarbonate buffer to a volume of 50 µl/well
and incubated the plates overnight at 4°C. We then
blocked the plates with 5% bovine serum albumin in
PBS, 200 µl/well, and incubated for 30 minutes at
room temperature.

Subsequently, we added 50 µl of sample or
standard to each well, incubating for 4 hours at
room temperature. We then added 50 µl/well of OKT4A
at 0.1 µg/ml and incubated overnight at 4°C. Using
a Hyclone Kit (Hyclone) we then carried out the fol-
lowing steps. First, we added 1 drop of rabbit
anti-mouse IgG2a to each well and incubated the
plates for 1 hour at room temperature. We then added
100 µl of peroxidase-labeled anti-rabbit IgG, diluted
1:4000 with 5% BSA/PBS to each well, and incubated
for 1 hour at room temperature.

We prepared a substrate reagent as follows.
W diluted substrate reagent 1:10 in distilled water
and added tw O-phenyl- thyl n -diamine ("OPD")
chromophore tablets p r 10 ml of substrate. W l t

the mixture dissolve thoroughly by mixing with a
vortex. Alternatively, a TMB peroxidase substrate
system (Kirkegaard & Perry Catalogue #50-76-00) may
be used. Subsequently, we added 100 μl of the

5   chromophore solution to each well, incubated for
10-15 minutes at room temperature and then stopped
the color development with 100 μl of 1N H_2SO_4. We
then measured OD at 490 nm, using an ELISA plate
reader.

10   The results of the assay are demonstrated
in the tables below.

-89-

Monkey #7-91

rsT4 Level
(ng/ml)

Time(hr)	Day 1	Day 7	Day 14	Day 28
0	22.7*	96.5	158.0	19.8
1	278.8	199.6	360.7	238.3
2	281.8	366.8	306.4	441.1
4	214.9	246.6	363.9	393.2
5				290.4
8	72.3	105.0	199.4	
9**	246.2			
10	259.6			
12	136.0			
22	23.8			
24	13.4			

Monkey #7-92

rsT4 Level
(ng/ml)

Time(hr)	Day 1	Day 7	Day 14	Day 28
0	6.7*	56.0	106.3	60.9
1	87.2	225.8	178.0	437.7
2	254.2	377.9	253.2	770.6
4	170.0	167.3	308.2	821.5
5				898.3
8	118.9	101.2	176.5	
9**	405.1			
10	523.5			
12	371.5			
22	48.4			
24	39.4			

* - background

** - second injection administered after the
collection of the 8 hour sample.

Polyvalent Forms Of Recombinant Soluble T4

Receptors may be characterized by their
affinity for specific ligands, such that, at equili-
brium, the intrinsic affinity (K_a) between monovalent
receptor and monovalent ligand can be defined as
$[RL]/[R_f][L_f]$, where [RL] is the concentration of
receptor (R) bound to ligand (L) and $[R_f]$ and $[L_f]$
are the concentrations of free receptor and ligand,

respectively [P. A. Underwood, in _Advances In Virus Research_, ed. K. Maramorosch et al., 34, pp. 283-309 (1988)].

For a polyvalent receptor (with a valency of n) binding to a polyvalent ligand (with a valency of m), a functional affinity can be defined as $n[R_b]/n[R_f]m[L_f]$, where $[R_b]$ is the concentration of bound receptor sites, and $n[R_f]$ and $m[L_f]$ are, respectively, the concentrations of free receptor and ligand binding sites. The effect of increasing the valence (the number of binding sites) is to enhance the stability of ligand-receptor complexes. The affinity of a polyvalent receptor for a polyvalent ligand will depend on three factors: the intrinsic association constant of each binding site, the valency (number of binding sites) and the topicological relationship between the receptor and ligand binding sites. Under some circumstances, polyvalent binding interactions will lead to higher functional affinity. The decreased dissociation rate of polyvalent ligands with polyvalent receptors results in an increased functional affinity [C. L. Hornick and F. Karush, _Immunochemistry_, 9, pp. 325-40 (1972); I. Otterness and F. Karush, "Principles Of Antibody Reactions", in _Antibody As A Tool_, ed. J. J. Marchalonais and G.W. Warr, pp. 97-137 (1982)].

The simplest case for receptor polyvalency increasing functional affinity is represented by a bivalent soluble receptor, such as an antibody molecule, which has two identical ligand binding sites, each capable of independently binding antigen with equal affinity. If the antigen is displayed polyvalently, for example, chemically coupled to a solid support such that the spacing between antigenic sites can be bridged by the antibody's two antigen binding arms, the functional affinity of the antibody for the antigen coupled to th solid support would b

greater than the intrinsic affinity of the antibody binding site for the monovalent antigen [D. Crothers and H. Metzger, _Immunochemistry_, 9, pp. 341-57 (1972)]. Because virus particles represent poly-valent antigens, the greater functional affinity of antibodies for polyvalent antigens is an important factor for antibody-directed virus neutralization.

The association of recombinant soluble T4 and the HIV major envelope glycoprotein gp120 is an example of monovalent receptor binding to monovalent ligand. The affinity of this interaction has been measured, and the association between T4 and gp120 has a dissociation constant $K_d = 4 \times 10^{-9}$ M [L. Lasky et al., _Cell_, 50, pp. 975-88 (1987)].

Using the antibody analogy, we believe that polyvalent rsT4 will demonstrate a greater affinity for HIV-infected cells displaying gp120 than monovalent rsT4 and the topicological relation-ship between gp120 on the virus particle or the infected cell surface, will determine the degree to which polyvalent rsT4 exhibits higher functional affinity than monovalent rsT4. One example of a polyvalent rsT4 is described below, with respect to the production of a recombinant bivalent rsT4 con-sisting of two tandem repeats of amino acids 3-178, followed by the C-terminal 199 amino acids of rsT4.3. According to this invention, a "polyvalent" receptor possesses two or more binding sites for a given ligand. Furthermore, the intrinsic affinity of each ligand binding site of a given polyvalent receptor need not be identical.

As shown in Figure 42, to construct bivalent rsT4, we digested pBG391 with _Nhe_I, which cleaves after the valine at position 178 in rsT4, and removed the _Nhe_I 5' overhang with mung bean nuclease. Next, w cleaved with _Bgl_II to remove th C-t rminal half of th rsT4 coding sequence in pBG391. Finally, we

ligat d a <u>DraI</u>-<u>Bal</u>II fragment containing the coding sequence for rsT4 amino acids 3 (lysine) through 377 (isoleucine) to the cleaved pBG391 to create pBiv.1, a plasmid coding for a fusion protein with a tandem

5   duplication of the N-terminal 176 amino acids of rsT4, followed by the C-terminal 199 amino acids of rsT4.3. The protein produced by this plasmid, therefore, contains two adjacent N-terminal gp120-binding or OKT4A-binding domains (defined by amino

10 acid residues 3 through 111 of rsT4.111), followed by one OKT4-binding C-terminal domain (Figure 43).

        pBiv.1 was transfected by electroporation into COS 7 cells to test expression of the bivalent rsT4 protein. Three days later, we tested the con-

15 ditioned medium of the transfected cells for the presence of the rsT4 bivalent protein by immuno-precipitation, followed by Western blot analysis of the precipitated protein. Both OKT4A and OKT4 were used for immuno-precipitation to determine that the

20 OKT4 epitope and at least one of the OKT4A epitopes had folded correctly. Both antibodies precipitated a protein of the predicted apparent molecular weight (60,000d) from the conditioned medium of the cells.

        Bivalent rsT4 may be purified by immuno-

25 affinity purification from an OKT4 column and the purified protein may then be used to perform quanti-tative competition assays with rsT4.3. We believe that the bivalent molecule would demonstrate equi-valent competition against rsT4.3 for OKT4 binding,

30 but significantly greater competition against mono-valent rsT4 for OKT4A binding. The ability of bivalent recombinant soluble T4 to block syncytium formation may also be demonstrated in the C8166 fusion assay. We also believe that bivalent

35 r combinant soluble T4 would block syncytium formation at significantly lower conc ntrations than monovalent rsT4; based up n the high r

-93-

functional affinity of bivalent recombinant
soluble T4 for gp120.

        According to alternate embodiments of this
invention, other methods for producing polyvalent
5   rsT4 may be employed. For example, polyvalent rsT4
may be produced by chemically coupling rsT4 to any
clinically acceptable carrier molecule, a polymer
selected from the group consisting of Ficoll, poly-
ethylene glycol or dextran, using conventional
10 coupling techniques. Alternatively, rsT4 may be
chemically coupled to biotin, and the biotin-rsT4
conjugate then allowed to bind to avidin, resulting
in tetravalent avidin/biotin/rsT4 molecules. And
rsT4 may be covalently coupled to dinitrophenol
15 (DNP) or trinitrophenol (TNP) and the resulting
conjugate precipitated with anti-DNP or anti-TNP-
IgM, to form decameric conjugates with a valency of
10 for rsT4 binding sites.

        Alternatively, a recombinant chimeric
20 antibody molecule with rsT4 sequences substituted
for the variable domains of either or both of the
immunoglobulin molecule heavy and light chains may
be produced. Because recombinant soluble T4
possesses gp120 binding activity, the construction
25 of a chimeric antibody having two soluble T4 domains
and having unmodified constant region domains could
serve as a mediator of targeted killing of HIV-
infected cells that express gp120.

        For example, chimeric rsT4/IgG$_1$ may be
30 produced from two chimeric genes -- an rsT4/human
kappa light chain chimera (rsT4/C$_{kappa}$) and an
rsT4/human gamma 1 heavy chain chimera
(rsT4/C$_{gamma-1}$). Both C$_{kappa}$ and C$_{gamma-1}$ regions
have been isolated from human recombinant DNA
35 libraries, and each has been subcloned into animal
cell s l cti n vectors containing either the
bacterial neo resistance or bacterial gpt markers

for selection in animal cell hosts against the
antibiotic G418 or mycophenolic acid, respectively.

To construct rsT4/$C_{gamma-1}$ and rsT4/C_{kappa}
chimeric genes, an rsT4 gene segment, including at
least the secretory signal sequence and the N-terminal
110 amino acid residues of the mature rsT4 coding
sequence and including a splice donor or portion
thereof, is placed upstream of the gamma-1 and kappa
constant domain exons. A suitable restriction
enzyme may be used to cut within the intron down-
stream of the desired rsT4 coding sequence, thus
providing a donor splice site. Subsequently, a
suitable restriction enzyme is used to cut within
the introns upstream of the kappa and gamma-1
coding regions. The rsT4 sequence is then joined to
the kappa or gamma-1 constant region sequence, such
that the rsT4 intron sequence is contiguous with the
gamma-1 and kappa introns. In this way, an acceptor
splice site is provided by the kappa or gamma-1
constant region intron. Alternatively, rsT4 chimeric
genes may be constructed without the use of introns,
by fusing a suitable rsT4 cDNA gene segment directly
to the gamma-1 or kappa coding regions.

The rsT4/$C_{gamma-1}$ and rsT4/C_{kappa} vectors
may then be cotransfected, for example, by electro-
poration into lymphoid or non-lymphoid host cells.
Following transcription and translation of the two
chimeric genes, the gene products may assemble into
chimeric antibody molecules.

Expression of the chimeric gene products
may be measured by an enzyme-linked immunoadsorbant
assay (ELISA) that utilizes monoclonal anti-T4 anti-
body OKT4A, as described infra, or in gp120 competi-
tion assays and radioimmunoassays, as described infra.
Activity f the rsT4/IgG_1 chim ras may be measured
by incubating them with HIV-inf cted cells in the
presence of human complement, followed by quantitating

subsequent complement-mediated lysis of these cells.
Alternatively, activity may be measured in HIV repli-
cation and HIV syncytium assays as described infra.

In order to determine if bivalent rsT4 has
a greater potency than monovalent rsT4, we mixed
OKT4, at various concentrations, together with a
constant concentration of rsT4, so that the molar
ratio of OKT4:rsT4 varied between 0.2 and 4. After
preincubating the mixture overnight at 4°C, we added
aliquots to the HIV syncytium assay described infra.
OKT4 has no observable effect in this assay when used
alone. In addition, the concentration of recombinant
soluble T4 chosen did not cause inhibition in this
assay. Accordingly, we looked for indications that
the OKT4/rsT4 mixture was more potent than rsT4 alone.
We observed that at ratios of OKT4:rsT4 greater than
0.2, partial to complete inhibition of syncytium
formation occurred. We believe that under conditions
where two rsT4 molecules are bound to 1 OKT4 molecule,
the greatest inhibitory effect should be found.

Thus, polyvalent, as well as monovalent
forms of recombinant soluble T4 are useful in the
compositions and methods of this invention.

Microorganisms and recombinant DNA mole-
cules prepared by the processes of this invention
are exemplified by cultures deposited in the In Vitro
International, Inc. culture collection, in Linthicum,
Maryland, on September 2, 1987, and identified as:

BG378:   E.coli MC1061/pBG378
199-7:   E.coli MC1061/p199-7
170-2:   E.coli JA221/p170-2
EC100:   E.coli JM83/pEC100
BG377:   E.coli MC1061/pBG377
BG380:   E.coli MC1061/pBG380
BG381:   E.coli MC1061/pBG381

These cultures were assigned accession
numbers IVI 10143-10149, respectively.

In addition, microorganisms and recombinant DNA molecules according to this invention are exemplified by cultures deposited in the In Vitro International, Inc. culture collection, in Linthicum, Maryland, on January 6, 1988, and identified as:

BG-391: E.coli MC1061/pBG391

BG-392: E.coli MC1061/pBG392

BG-393: E.coli MC1061/pBG393

BG-394: E.coli MC1061/pBG394

BG-396: E.coli MC1061/pBG396

203-5 : E.coli SG936/p203-5.

These cultures were assigned accession numbers IVI 10151-10156, respectively.

Microorganisms and recombinant DNA molecules according to this invention are also exemplified by cultures deposited in the In Vitro International, Inc. culture collection, in Linthicum, Maryland, on August 24, 1988 and identified as:

211-11: E.coli A89/pBG211-11

214-10: E.coli A89/pBG214-10

215-7 : E.coli A89/pBG215-7

These cultures were assigned accession numbers IVI 10183-10185 respectively.

While we have hereinbefore described a number of embodiments of this invention, it is apparent that our basic constructions can be altered to provide othe embodiments which utilize the processes and compositions of this invention. Therefore, it will be appreciated that the scope of this invention is to be defined by the claims appended hereto rather than by the specific embodiments which have been presented hereinbefore by way of example.

CLAIMS

We claim:

1.   A DNA sequence selected from the group consisting of:

5   (a)   the DNA inserts of p199-7, pBG377, pBG380, pBG381, p203-5, pBG391, pBG392, pBG393, pBG394, pBG395, pBG396, pBG397, p211-11, p214-10 and p215-7;

              (b)   DNA sequences which hybridize to one or more of the foregoing DNA inserts and which

10   code on expression for a soluble T4-like polypeptide; and

              (c)   DNA sequences which code on expression for a soluble T4-like polypeptide coded for on expression by any of the foregoing DNA inserts

15   and sequences.

              2.   The DNA sequence according to claim 1, wherein said DNA sequence (b) codes on expression for a soluble T4-like polypeptide which inhibits adhesion between T4$^+$ lymphocytes and infective agents

20   which target T4$^+$ lymphocytes and which inhibits interaction between T4$^+$ lymphocytes and antigen presenting cells and targets of T4$^+$ lymphocyte mediated killing.

              3.   A recombinant DNA molecule comprising

25   a DNA sequence selected from the group consisting of the DNA sequences of claim 1 or 2, said DNA sequence being operatively linked to an expression control sequence in said recombinant DNA molecule.

              4.   The recombinant DNA molecule according

30   to claim 3, wherein said expression control sequence is selected from the group consisting of the early or late promoters of SV40 or aden virus, the lac system, the trp system, the TAC system, the TRC

system, the major operator and promoter regions of
phage λ, the control regions of fd coat protein, the
promoter for 3-phosphoglycerate kinase or other
glycolytic enzymes, the promoters of acid phosphatase,
5 the polyhedron promoter of the baculovirus system
and the promoters of the yeast α-mating factors.

5.   A unicellular host transformed with a
recombinant DNA molecule selected from the group
consisting of the recombinant DNA molecules of claim 3
10 or 4.

6.   The host according to claim 5, wherein
said host is selected from the group consisting of
strains of E.coli, Pseudomonas, Bacillus,
Streptomyces, fungi, animal cells, plant cells,
15 insect cells and human cells in tissue culture.

7.   A polypeptide coded for on expression
by a DNA sequence selected from the group consisting
of the DNA sequences of claim 1 or 2, said polypep-
tide being essentially free of other proteins of
20 human origin.

8.   The polypeptide according to claim 7,
wherein said polypeptide is selected from the group
consisting of a polypeptide of the formula AA_{-23}-AA_{362}
of Figure 3, a polypeptide of the formula AA_{1-362} of
25 Figure 3, a polypeptide of the formula Met-AA_{1-362}
of Figure 3, a polypeptide of the formula AA_{1-374} of
Figure 3, a polypeptide of the formula Met-AA_{1-374}
of Figure 3, a polypeptide of the formula AA_{1-377} of
Figure 3, a polypeptide of the formula Met-AA_{1-377}
30 of Figure 3, a polypeptide of the formula AA_{-23}-AA_{374}
of Figure 3, a polypeptide of the formula AA_{-23}-AA_{377}
of Figure 3.

9.   The polypeptide according to claim 7, wherein said polypeptide is selected from the group consisting of a polypeptide of the formula AA_{-23}-AA_{182} of Figure 16, a polypeptide of the formula AA_1-AA_{182} of Figure 16, a polypeptide of the formula Met-AA_{1-182} of Figure 16, a polypeptide of the formula AA_{-23}-AA_{182} of Figure 16, followed by the amino acids asparagine-leucine-glutamine-histidine-serine-leucine, a polypeptide of the formula AA_1-AA_{182} of Figure 16, followed by the amino acids asparagine-leucine-glutamine-histidine-serine-leucine, a polypeptide of the formula Met-AA_{1-182} of Figure 16, followed by the amino acids asparagine-leucine-glutamine-histidine-serine-leucine, a polypeptide of the formula AA_{-23}-AA_{113} of Figure 16, a polypeptide of the formula AA_1-AA_{113} of Figure 16, a polypeptide of the formula Met-AA_{1-113} of Figure 16, a polypeptide of the formula AA_{-23}-AA_{111} of Figure 16, a polypeptide of the formula AA_1-AA_{111} of Figure 16, a polypeptide of the formula Met-AA_{1-111} of Figure 16, a polypeptide of the formula AA_{-23}-AA_{131} of Figure 16, a polypeptide of the formula AA_1-AA_{131} of Figure 16, a polypeptide of the formula Met-AA_{1-131} of Figure 16, a polypeptide of the formula AA_{-23}-AA_{145} of Figure 16, a polypeptide of the formula AA_1-AA_{145} of Figure 16, a polypeptide of the formula Met-AA_{1-145} of Figure 16, a polypeptide of the formula AA_{-23}-AA_{166} of Figure 16, a polypeptide of the formula AA_1-AA_{166} of Figure 16, a polypeptide of the formula Met-AA_{1-166} of Figure 16, or portions thereof.

10.   The polypeptide according to claim 7, wherein said polypeptide is selected from the group consisting of a polyp ptide of the formula AA_{-23}-AA_{362} of mature T4 protein, a polypeptide of the formula AA_{1-362} of mature T4 protein, a p lypeptide of the formula M t-AA_{1-362} of mature T4 protein, a polypep-

-100-

tide of the formula AA_{1-374} of mature T4 protein, a
polypeptide of the formula Met-AA_{1-374} of mature T4
protein, a polypeptide of the formula AA_{1-377} of
mature T4 protein, a polypeptide of the formula
5   Met-AA_{1-377} of mature T4 protein, a polypeptide of
the formula AA_{-23}-AA_{374} of mature T4 protein, a poly-
peptide of the formula AA_{-23}-AA_{377} of mature T4 pro-
tein, or portions thereof.

           11.   The polypeptide according to claim 7,
10   wherein said polypeptide is selected from the group
consisting of a polypeptide of the formula AA_{-23}-AA_{182}
of mature T4 protein, a polypeptide of the formula
AA_{1}-AA_{182} of mature T4 protein, a polypeptide of the
formula Met-AA_{1-182} of mature T4 protein, a polypep-
15   tide of the formula AA_{-23}-AA_{182} of mature T4 protein,
followed by the amino acids asparagine-leucine-
glutamine-histidine-serine-leucine, a polypeptide of
the formula AA_{1}-AA_{182} of mature T4 protein, followed
by the amino acids asparagine-leucine-glutamine-
20   histidine-serine-leucine, a polypeptide of the formula
Met-AA_{1-182} of mature T4 protein, followed by the
amino acids asparagine-leucine-glutamine-histidine-
serine-leucine, a polypeptide of the formula
AA_{-23}-AA_{113} of mature T4 protein, a polypeptide of
25   the formula AA_{1}-AA_{113} of mature T4 protein, a polypep-
tide of the formula Met-AA_{1-113} of mature T4 protein,
a polypeptide of the formula AA_{-23}-AA_{111} of mature
T4 protein, a polypeptide of the formula AA_{1}-AA_{111}
of mature T4 protein, a polypeptide of the formula
30   Met-AA_{1-111} of mature T4 protein, a polypeptide of
the formula AA_{-23}-AA_{131} of mature T4 protein, a poly-
peptide of the formula AA_{1}-AA_{131} of mature T4 protein,
a polypeptide of the formula Met-AA_{1-131} of mature
T4 pr tein, a polypeptide of the formula AA_{-23}-AA_{145}
35   f mature T4 pr tein, a polypeptid f the formula
AA_{1}-AA_{145} of mature T4 protein, a polyp ptid of th

-101-

formula Met-AA_{1-145} of mature T4 protein, a polypeptide of the formula $AA_{-23}-AA_{166}$ of mature T4 protein, a polypeptide of the formula AA_1-AA_{166} of mature T4 protein, a polypeptide of the formula Met-AA_{1-166} of

5   mature T4 protein, or portions thereof.

12. A method for producing a polypeptide selected from the group consisting of the polypeptides of any one of claims 7 to 11 comprising the step of culturing a unicellular host transformed with a recom-

10   binant DNA molecule selected from the group consisting of the recombinant DNA molecules of claim 3 or 4.

13. A pharmaceutical composition comprising an immunotherapeutic or immunosuppressive effective amount of a polypeptide selected from the group con-

15   sisting of the polypeptides of any one of claims 7 to 11 and a pharmaceutically acceptable carrier.

14. A method for treating patients comprising the step of treating them in a pharmaceutically acceptable manner with a composition selected

20   from the group consisting of the composition of claim 13.

15. The method according to claim 14, wherein the patient is treated by intramuscular injection of the composition.

25   16. A diagnostic composition for detecting or for monitoring the course of HIV infection comprising a diagnostic effective amount of a polypeptide selected from the group consisting of the polypeptides of any one of claims 7 to 11.

30   17. A method for detecting or for monitoring th course of HIV inf ction comprising the

step of employing as a diagnostic a composition
selected from the group consisting of the compositions
of claim 16.

18.   A means for detecting or for monitoring
the course of HIV infection comprising a composition
selected from the group consisting of the compositions
of claim 16.

19.   A pharmaceutical composition compris-
ing an immunotherapeutic or immunosuppressive effec-
tive amount of antibody to a polypeptide selected
from the group consisting of the polypeptides of any
one of claims 7 to 11 and a pharmaceutically accept-
able carrier.

20.   A method for treating patients com-
prising the step of treating them in a pharmaceuti-
cally acceptable manner with a composition according
to claim 19.

21.   The use of a polypeptide selected
from the group consisting of the polypeptides of any
one of claims 7 to 11 to purify HIV virus.

22.   The use according to claim 20, wherein
the HIV virus is purified from a biological sample.

23.   A method for purifying HIV virus from
a sample comprising the step of exposing the sample
to a polypeptide selected from the group consisting
of the polypeptides of any one of claims 7 to 11.

24.   The method according to claim 22,
wherein the sample is a biological sample.

25. A DNA sequence comprising the DNA insert of p170-2, said sequence coding on expression for a T4-like polypeptide.

26. A recombinant DNA molecule comprising a DNA sequence selected from the group consisting of the DNA sequence of claim 25, said DNA sequence being operatively linked to an expression control sequence in said recombinant DNA molecule.

27. A unicellular host transformed with a recombinant DNA molecule according to claim 26.

28. A polypeptide coded for on expression by a DNA sequence of claim 25, said polypeptide being essentially free of other proteins of human origin.

29. A pharmaceutical composition comprising an immunotherapeutic or immunosuppressive amount of a soluble protein receptor and a pharmaceutically acceptable carrier.

30. A method for treating patients comprising the step of treating them in a pharmaceutically acceptable manner with a pharmaceutical composition of claim 29.

31. A diagnostic composition for detecting or for monitoring the course of viral infection comprising a diagnostic effective amount of a soluble protein receptor.

32. A method for detecting or for monitoring the cours of a viral inf ction comprising th step f employing as a diagnostic a diagnostic effectiv amount f a soluble pr tein recept r.

-104-

33.   A means for detecting or for moni-
toring the course of a viral infection comprising
a soluble protein receptor.

34.   A DNA sequence selected from the group
consisting of:

(a)   the DNA insert of pBiv.1;

(b)   DNA sequences which hybridize to
the DNA insert of pBiv.1 and which code on expression
for a polyvalent soluble T4-like polypeptide; and

(c)   DNA sequences which code on
expression for a polyvalent soluble T4-like polypep-
tide coded for by the DNA insert of pBiv.1.

35.   A recombinant DNA molecule comprising
a DNA sequence selected from the group consisting of
the DNA sequences of claim 34, said DNA sequence
being operatively linked to an expression control
sequence in said recombinant DNA molecule.

36.   A unicellular host transformed with a
recombinant DNA molecule according to claim 35.

37.   A polypeptide coded for on expression
by a DNA sequence selected from the group consisting
of the DNA sequences according to claim 34, said
polypeptide being essentially free of other proteins
of human origin.

38.   The polypeptide according to claim 7,
wherein said polypeptide is polyvalent.

39.   A method for producing a polyvalent
polypeptide comprising the steps of:

(a)   culturing a unicellular host
transformed with a recombinant DNA molecule according
to claim 3 or 4 to produce a polypeptide; and

(b) coupling said polypeptide to a
carrier to form a polyvalent polypeptide.

40. A DNA sequence comprising:

(a) a first portion comprising a DNA
sequence coding for the constant region of an immuno-
globulin light chain; and

(b) a second portion comprising a
DNA sequence according to claim 1 or 2, or portions
thereof, said second portion being joined upstream
of said first portion.

41. A DNA sequence comprising:

(a) a first portion comprising a DNA
sequence coding for the constant region of an immuno-
globulin heavy chain; and

(b) a second portion comprising a
DNA sequence according to claim 1 or 2, or portions
thereof, said second portion being joined upstream
of said first portion.

42. An expression vector comprising the
DNA sequence according to claim 40.

43. An expression vector comprising the
DNA sequence according to claim 41.

44. An expression vector comprising the
DNA sequence according to claim 40 and the DNA
sequence according to claim 41.

45. A method for producing a chimeric
rsT4/IgG$_1$ comprising the step of co-transfecting a
host cell with the expression vector according to
claim 42 and the expression v ctor according to
claim 43.

46. A method for producing a chimeric rsT4/IgG$_1$ comprising the step of transfecting a host cell with the expression vector according to claim 44.

47. A chimeric rsT4/IgG$_1$ produced by the method according to claim 45 or 46.

48. A pharmaceutical composition comprising an immunotherapeutic or immunosuppressive effective amount of a polypeptide according to claim 37 or 38.

49. A method for treating patients comprising the step of treating them in a pharmaceutically acceptable manner with a composition according to claim 48.

50. A diagnostic composition for detecting or for monitoring the course of HIV infection comprising a diagnostic effective amount of a polypeptide according to claim 37 or 38.

51. A pharmaceutical composition comprising an immunotherapeutic or immunosuppressive effective amount of a chimeric rsT4/IgG$_1$ according to claim 47.

52. A method for treating patients comprising the step of treating them in a pharmaceutically acceptable manner with a composition according to claim 51.

1/93

FIG. 1

FIG. 2

FIG. 3

```
H                      P                                    BBE S
i                      s                              M M NBscNc
n                      t                              n n laptosr
d         REX5'        I                              l l anlNRpP
I                                                     : : 422122:
                                                            /////
....-....TTTAAGCACGACTCTGCAGAAGGAACAAAGCACCCTCCCCACTGGGCTCCTGG ---------- 60
:--------------------------+---------+---------+---------+---------
AAAAAAAAAATTCGTGCTGAGACGTCTTCCTTGTTTCGTGGGAGGGGTGACCCGAGGACC

   F   F   L   S   T   T   L   Q   K   E   Q   S   T   L   P   T   G   L   L   V -


            BH                                       F N                  B   BM
     A   BsgNSS                      M   M           n s                  b   bn
     l   apisas                      n   n           u p                  v   vl
     u   nlApct                      l   l           4 B                  1   11
     l   221211                      l   l           H 2
          /////
     TTGCAGAGCTCCAAGTCCTCACACAGATACGCCTGTTTGAGAAGCAGCGGGCAAGAAAGA ---------- 120
61 --------------------------+---------+---------+---------+---------+
     AACGTCTCGAGGTTCAGGAGTGTGTCTATGCGGACAAACTCTTCGTCGCCCGTTCTTTCT

       A   E   L   Q   V   L   T   Q   I   R   L   P   E   K   Q   R   A   R   K   T -


                   S                B           B       P S              D   S
     M       H   HDHaP              s           DBsNADNpPaS              D   DHNPa
     n       g   graus              t           dapsvrlusui              d   ralsu
     l       a   aae9s              X           enlpaaaMs9n              e   aeas9
     l       l   12361              l           12222241161              l   23416
                    /                                 /// / ////               /
     CGCAAGCCCAGAGGCCCTGCCATTTCTGTGGGCTCAGGTCCCTACTGGCTCAGGCCCCTG ---------- 180
121 --------------------------+---------+---------+---------+---------+
     GCGTTCGGGTCTCCGGGACGGTAAAGACACCCGAGTCCAGGGATGACCGAGTCCGGGGAC

       Q   A   Q   R   P   C   H   F   C   G   L   R   S   L   L   A   Q   A   P   A -

                   REX Splice
                   Donor
              M M  MHM           S H                        B   B
              n n  nan         HDMNc i                      b   b
              l l  lel         pscr n                       v   v
              1 1  131         apiF f                       1   1
                              2111 1
                        Met        ///
     CCTCCCTCGGCAAGGCCACA[ATG]AACCGGGGAGTCCCTTTTAGGCACTTGCTTCTGGTGC ---------- 240
181 --------------------------+---------+---------+---------+---------+
     GGAGGGAGCCGTTCCGGTGTTACTTGGCCCCTCAGGGAAAATCCGTGAACGAAGACCACG

       S   L   G   K   A   T   M   N   R   G   V   P   F   R   H   L   L   L   V   L -
                               AA-23
       H                       F
       i        HH           nM M   D        B   B
       n        ha           un n   d        b   b
       P        ae           41 l   e        v   v
       l        12           H1 l   l        1   1
     TGCAACTGGCGCTCCTCCCCAGCAGCCACTCAGGGAAAGAAAGTGGTGCTGGGCAAAAAG ---------- 300
241 --------------------------+---------+---------+---------+---------+
     ACGTTGACCGCGAGGAGGGGTCGTCGGTGAGTCCCTTTCTTTCACCACGACCCGTTTTTTC

       Q   L   A   L   L   P   A   A   T   Q   G   K   R   V   V   L   G   K   K   G -

                               R        A                    M   MM  M
                               s        l       AAₗ          b   bb  b
                               a        u                    o
                               l        l                    2   22  2
     GGGATACAgtggaactgacctgtacagcttcccagaagaagagcatacaattccactgga ---------- 360
301 --------------------------+---------+---------+---------+---------+
     CCCTATGTCACCTTGACTGGACATGTCGAAGGGTCTTCTTCTCGTATGTTAAGGTGACCT

       D   T   V   E   L   T   C   T   A   S   Q   K   K   S   I   Q   F   H   W   K -
```

4/93

FIG. 3(cont'd)

```
                  H                          B      P    P           S
                  i                          NBsN   o    o           AaS
                  n                          laps   k    k           vui
                  f                          anlp   1    1           a9n
                  l                          4222                    261
                                              //                 |    //
      AAAACTCCAACCAGATAAAGATTCTGGGAAATCAGGGCTCCTTCTTAACTAAAGGTCCAT
  361 --------------------------------+---------+---------+------------ 420
      TTTTGAGGTTGGTCTATTTCTAAGACCCTTTAGTCCCGAGGAAGAATTGATTTCCAGGTA

        N  S  N  Q  I  K  I  L  G  N  Q  G  S  P  L  T  K  G  P  S -

                  S  FH      H                        S        REX
      A         MNDa niHT    i                  M    MANaSS     Splice
      l         bdpu unhb    n                  b    bvluit     Acceptor
      u         oen3 DPaa    f                  o    oaa9ny
      1         121A 2111    1                  2    224611
                /  /  /  /                            ///
      CCAAGCTGAATGATCGCGCTGACTCAAGAAGAAGCCTTCGGGACCAAGGAAACTTTCCCC
  421 --------------------------------+---------+---------+------------ 480
      GGTTCGACTTACTAGCGCGACTGAGTTCTTCTTCGGAAGCCCTGGTTCCTTTGAAAGGGG

        K  L  N  D  R  A  D  S  R  R  S  L  R  D  Q  G  N  F  P  L -

                  S      H                H                            S
      BMDNa       i      A                i  D   M    M      M  M      MAMMaM
      cbpdu       n      f                n  d   b    b      n  n      nvnnun
      lone3       f      l                f  e   o    o      l  l      lal191
      1112A       1      2                1  1   2    2      1  1      121161
      /  ///                              /
      TGATCATCAAGAATCTTAAGATAGAAGACTCAGATACTTACATCTGTGAAGTGGAGGACC
  481 --------------------------------+---------+---------+------------ 540
      ACTAGTAGTTCTTAGAATTCTATCTTCTGAGTCTATGAATGTAGACACTTCACCTCCTGG

        I  I  K  N  L  K  I  E  D  S  D  T  Y  I  C  E  V  E  D  Q -

      S                    M                            E
      i                    a                            c
      n                    e                            o
      l                    l                            R
      /                    /
      AGAAGGAGGAGGTGCAATTGCTAGTGTTCGGATTGACTGCCAACTCTGACACCCACCTGC
  541 --------------------------------+---------+---------+------------ 600
      TCTTCCTCCTCCACGTTAACGATCACAAGCCTAACTGACGGTTGAGACTGTGGGTGGACG

        K  E  E  V  Q  L  L  V  F  G  L  T  A  N  S  D  T  H  L  L -

      B     B                           BBE  S
      s     s                    S      BsscNc         D
      P     P                    t      aptosr         d
      M     M                    y      nlNRpF         e
      1     1                    1      221221         1
                                        /  ///
      TTCAGGGGCAGAGCCTGACCCTGACCTTGGAGAGCCCCCCTGGTAGTAGCCCCTCAGTGC
  601 --------------------------------+---------+---------+------------ 660
      AAGTCCCCGTCTCGGACTGGGACTGGAACCTCTCGGGGGGACCATCATCGGGGAGTCACG

        Q  G  Q  S  L  T  L  T  L  E  S  P  P  G  S  S  P  S  V  Q -

                   Hin                                              N
                   i                               M  M  M DM       Aa
      M  M         n      S                        b  b  n dn       lp
      n  n         f      t                           l el         uB
      1  1         1      y                        2  2  1 11       12
      1  1                1                                /          /
      AATGTAGGAGTCCAAGGGGTAAAAACATACAGGGGGGGGAAGACCCTCTCCGTGTCTCAGC
  661 --------------------------------+---------+---------+------------ 720
      TTACATCCTCAGGTTCCCCATTTTTGTATGTCCCCCCCCTTCTGGGAGAGGCACAGAGTCG

        r  R  S  P  R  G  K  N  I  Q  G  G  K  T  L  S  V  S  Q  L -
```

5/93

FIG. 3(cont'd)

```
      BBEH  SB        BES         N                              M
 P    A  BsscgNScsS B Nscc       Ns                             b
 v    l  aptoisarts a ltor       lp                             o
 u    u  nlNRApcFXt s aNRF       aH                             2
 2    1  2212121111 1 4121       31
 /       /////// /     /          /
    TGGAGCTCCAGGATAGTGGCACCTGGACATGCACTGTCTTGCAGAACCAGAAGAAGGTGG
721 --------+---------+---------+---------+---------+---------+ 780
    ACCTCGAGGTCCTATCACCGTGGACCTGTACGTGACAGAACGTCTTGGTCTTCTTCCACC
      E   L   Q   D   S   G   T   W   T   C   T   V   L   Q   N   Q   K   K   V   E  -
```

```
          M              NM  A        HS          M  MM M
          b              ha  l        at          n  nn n
          o              ee  u        eu          l  11 1
          2              11  1        31          1  11 1
    AGTTCAAAATAGACATCGTGGTGCTAGCTTTCCAGAAGGCCTCCAGCATAGTCTATAAGA
781 --------+---------+---------+---------+---------+---------+ 840
    TCAAGTTTTATCTGTAGCACCACGATCGAAAGGTCTTCCGGAGGTCGTATCAGATATTCT
      F   K   I   D   I   V   V   L   A   F   Q   K   A   S   S   I   V   Y   K   K  -
```

```
                                                        A
                                                        l
                                                        u
                                                        1
    AAGAGGGGGGAACAGGTGGAGTTCTCCTTCCCACTCGCCTTTACAGTTGAAAAGCTGACGG
841 --------+---------+---------+---------+---------+---------+ 900
    TTCTCCCCCTTGTCCACCTCAAGAGGAAGGGTGAGCGGAAATGTCAACTTTTCGACTGCC
      E   G   E   Q   V   E   F   S   F   P   L   A   F   T   V   E   K   L   T   G  -
```

```
                                                  P      S
          A              M  M                H  MHMMfMMMNDa
          l              n  n                p  npnnlnbdpu
          u              l  l                b  1h11Mloen3
          1              1  1                1  111111121A
                                                   /  //
    GCAGTGGCGAGCTGTGGTGGCAGGCGGAGAGGGCTTCCTCCTCCAAGTCTTGGATCACCT
901 --------+---------+---------+---------+---------+---------+ 960
    CGTCACCGCTCGACACCACCGTCCGCCTCTCCCGAAGGAGGAGGTTCAGAACCTAGTGGA
      S   G   E   L   W   W   Q   A   E   R   A   S   S   S   K   S   W   I   T   S  -
```

```
                                      B      BES    P  S
          M  M          M    M        sM     scc  ADNpPDaS A
          n  n          b    b        ta     tor  vrlusdui l
          l  l          o    o        Ee     NRF  aaaMse9n u
          1  1          2    2        23     121  22411161 1
                                       /      /     / ////
    CTGACCTGAAGAACAAGGAAGTGTCTGTAAAACGGGTTACCCAGGACCCTAAGCTCCAGA
961 --------+---------+---------+---------+---------+---------+ 1020
    GACTGGACTTCTTGTTCCTTCACAGACATTTTGCCCAATGGGTCCTGGGATTCGAGGTCT
      D   L   K   N   K   E   V   S   V   K   R   V   T   Q   D   P   K   L   Q   M  -
```

```
                                      BE S
          A  H        H        M  scMc HS      D        M  M H
          l  p        p        n  tonr at      d        n  n p
          u  h        h        l  NR1F eu      1        l  l h
          1  1        1        1  1211 31      1        1  1 1
                                      ///  /
    TGGGCAAGAAGCTCCCGCTCCACCTCACCCTGCCCCAGGCCTTGCCTCAGTATGCTGGCT
1021 --------+---------+---------+---------+---------+---------+ 1080
    ACCCGTTCTTCGAGGGCGAGGTGGAGTGGGACGGGGTCCGGAACGGAGTCATACGACCGA
      G   K   K   L   P   L   H   L   T   L   P   Q   A   L   P   Q   Y   A   G   S  -
```

SUBSTITUTE SHEET

6/93

FIG. 3(cont'd)

```
                                                                    S  S BES
                    BES      S                                      f  f scc
        H           scc   MHMa                                      a  a tor
        P           tor   nanu                                      N  N NRF
        h           NRF   1el9                                      1  1 121
        1           121   1316                                      /          /
     CTGGAAACCTCACCCTGGCCCTTGAAGCGAAAACAGGAAAGTTGCATCAGGAAGTGAACC
1081 --------------------------+---------+---------+---------+---------+ 1140
     GACCTTTGGAGTGGGACCGGGAACTTCGCTTTTGTCCTTTCAACGTAGTCCTTCACTTGG

       G   N   L   T   L   A   L   E   A   R   T   G   K   L   H   Q   E   V   N   L-


                    H        H   D        A                M   M                P S
                    p        p   d        l                n   n        ADNNpPa
                    h        h   e        u                l   l        vrllusu
                    1        1   1        1                1   1        aaaaMs9
                                                                       2244116
                                                                       //////
     TGGTGGTGATGAGAGCCACTCAGCTCCAGAAAAAATTTGACCTGTGAGGTGTGGGGGACCCA
1141 ---------+----+---------+---------+---------+---------+---------+ 1200
     ACCACCACTACTCTCGGTGAGTCGAGGTCTTTTTTAAACTGGACACTCCACACCCCTGGGT

        V   V   M   R   A   T   Q   L   Q   K   N   L   T   C   E   V   W   G   P   T-


        S    S                                           T
        i    f   Df  MAM        DE   A        M   M      a
        n    a   da  nln        ds   l        n   n      q
        1    N   eN  lul        ep   u        l   l      1
             1   11  111        11   1        1   1
        /        /        /
     CCTCCCCTAAGCTGATGCTGAGCTTGAAACTGGAGAACAAGGAGGCAAAGGTCTCGAAGC
1201 ---------+---------+---------+---------+---------+---------+ 1260
     GGAGGGGATTCGACTACGACTCGAACTTTGACCTCTTGTTCCTCCGTTTCCAGAGCTTCG

         S   P   R   L   M   L   S   L   K   L   E   N   K   E   A   V   S   K   R-


                         M   M            DM                    F   FD   M       H
                         n   n            ds                    o   od   a       i
                         l   l            et                    k   ke   e       n
                         1   1            12                    1   11   3       f
                                          /                                     1
     GGGAGAAGGCGGTGTGGGTGCTGAACCCTGAGGCGGGGATGTGGCAGTGTCTGCTGAGTG
1261 ---------+---------+---------+----+---------+---------+---------- 1320
     CCCTCTTCCGCCACACCCACGACTTGGGACTCCGCCCCTACACCGTCACAGACGACTCAC

       E   K   A   V   W   V   L   N   P   E   A   G   M   W   Q   C   L   L   S   D-


                    P S                H                         S            S
        A           ADpPaS             i                         ANaS         HDNC
        v           vrusui             n                         vlui         pecr
        a           aaMs9n             f                         aa9n         apiF
        1           221161             1                         2361         2111
                    //  //                                       ///          ///
     ACTCGGGACAGGTCCTGCTGGAATCCAACATCAAGGTTCTGCCCACATGGTCCACCCCGG
1321 ---------+---------+---------+---------+---------+---------+ 1380
     TGAGCCCTGTCCAGGACGACCTTAGGTTGTAGTTCCAAGACGGGTGTACCAGGTGGGGCC

        S   G   Q   V   L   L   E   S   N   I   K   V   L   P   T   W   S   T   P   V-


        F            S                       H
        n          aH  B   B   H    H        Ag      HDNH              M  M   M   MMM
        u          ua  b   b   g    g        hl     peaa               n  n   b   bba
        4          9e  v   v   a    a        aD     apee               l  l   o   o e
        H          63  1   1   1    1        21     2113               1  1   2   221
     TGCAGCCAATGGCCCTGATTGTGCTGGGGGGCGTCGCCGGCCTCCTGCTTTTTCATTGGGC
1381 --------------------------+---------+----+---------+---------+ 1440
     ACGTCGGTTACCGGGACTAACACGACCCCCCGCAGCGGCCGGAGGACGAAAAGTAACCCG

        Q   P   M   A   L   I   V   L   G   G   V   A   G   L   L   L   F   I   G   L-
```

7/93

FIG. 3(cont'd)

```
          S  S                                    H  H
          f  f          B N HMBNN              BAgHBH1NN                      D
 M        a  a          a l psaal              ahihbanal                     d
 b        N  N          n a apnea              naDaeePra                     e
 2        1  1          1 4 21114              121112114                     1
                          /  /            /   ////
    →
TAGGCATCTTCTTCTGTGTCAGGTGCCGGCACCGAAGGCGCCAAGCAGAGCGGATGTCTC
1441 ------------------------------------------------------------ 1500
ATCCGTAGAAGAAGACACAGTCCACGGCCGTGGCTTCCGCGGTTCGTCTCGCCTACAGAG

     G  I  F  F  C  V  R  C  R  H  R  R  R  Q  A  E  R  M  S  Q -
```

```
      S           H                           B    BB
 MNDaP    F    i        D           M M     MH  MMH Ms  ssN   HM   M M
 bdpuo    o    n        d           n n     bp  bbp bp  pps   ps   n n
 oen3k    k    f        e           l l     oh  ooh oM  1Mp   ap   l l
 121A1    1    1        1           1 1     21  221 21  212   21   1 1
   / /                                            //    /
AGATCAAGAGACTCCTCAGTGAGAAGAAGACCTGCCAGTGCCCTCACCGGTTTCAGAAGA
1501 ------------------------------------------------------------ 1560
TCTAGTTCTCTGAGGAGTCACTCTTCTTCTGGACGGTCACGGGAGTGGCCAAAGTCTTCT

     I  K  R  L  L  S  E  K  K  T  C  Q  C  P  H  R  F  Q  K  T -
```

```
              N                    BE S        S              F      BES
 MMNMMs    M  M         scHc     MNDaX        n        scc  BM  BM
 bnlnbp    n  n         toar     bdpuh        u        tor  bn  bn
 olaloH    l  l         NReF     oen3o        4        NRF  vl  vl
 213121    1  1         1231     121A2        H        121  11  11
   /   /                  ///      / //                   /
```

```
                        Stop                          REX
                                                     Stop
CATGTAGCCCCATT[TGA]GGCACGAGGCCAGGCAGATCCCACTTG|AGCCTCCCCAGGTGT
1561 ------------------------------------------------------------ 1620
GTACATCGGGGTAAACTCCGTGCTCCGGTCCGTCTAGGGTGAAC|TCGGAGGGGTCCACA

     C  S  P  I  *  G  T  R  P  G  R  S  H  L  Q  P  P  Q  V  S -
```

```
             F                S             SBES        B
             nT               AaS      MNDasccX HS  s      H  M
             ub               vui      bdputorh at  t      a  n
             Da               a9n      oen3NRFo eu  X      e  l
             21               261      121A1212 31  1      3  1
              /                //          / / / /
CTGCCCCGCGTTTCCTGCCTGCGGACCAGATGAATGTAGCAGATCCCAGGCCTCTGGCCT
1621 ------------------------------------------------------------ 1680
GACGGGGCGCAAAGGACGGACGCCTGGTCTACTTACATCGTCTAGGGTCCGGAGACCGGA

     A  P  R  F  L  P  A  D  Q  M  N  V  A  D  P  R  P  L  A  S -
```

```
                                                BES              SS
                                                scc        DHNPHMNac
 M        M M          M MM M                   tor        ralspscur
 n        n n          n nn n                   NRF        aeasap19F
 1        1 1          1 11 1                    121       234121161
 1        1 1          1 11 1                     /          ////
CCTGTTCGCCTCCTCTACAATTTGCCATTGTTTCTCCTGGGTTAGGCCCCGGCTTCACTG
1681 ------------------------------------------------------------ 1740
GGACAAGCGGAGGAGATGTTAAACGGTAACAAAGAGGACCCAATCCGGGGCCGAAGTGAC

     C  S  P  P  L  Q  F  A  I  V  S  F  P  G  L  G  P  G  F  T  G -
```

```
               MMM                                     M M
               naB                                     n n
               1el                                     1 1
               111                                     1 1
GTTGAGTGTTGCTCTCTAGTTTCCAGAGGCTTAATCACACCGTCCTCCACGCCATTTCCT
1741 ------------------------------------------------------------ 1800
CAACTCACAACGAGAGATCAAAGGTCTCCGAATTAGTGTGGCAGGAGGTGCGGTAAAGGA

      *  V  L  L  S  S  S  P  Q  R  L  N  H  T  V  L  H  A  I  S  F -
```

8/93

FIG. 3(cont'd)

```
                    M                                        M M
                    a                                        n n
                    e                                        l l
                    1                                        1 1

        TTTCCTTCAAGCCTAGCCCTTCTCTCATTATTTCTCTCTGACCCTCTCCCCACTGCTCAT
1801    ------------------------------------------------------------  1860
        AAAGGAAGTTCGGATCGGGAAGAGAGTAATAAAGAGAGACTGGGAGAGGGGTGACGAGTA

         S  P  K  P  S  P  S  L  I  I  S  L  *  P  S  P  H  C  S  P -

        B     BE SS                S      BE S                  H      H
        aMDNscNacX                 a   H   scMCM      M M       P      P
        mbpdtolurb                 u   a   tonrn      n n       h      h
        HoneNRa3Fo                 9   e   NR1F1      l l       1      1
        1112124A12                 6   3   12111      1 1
        / / /////                  / /  /              /
        TTGGATCCCAGGGGAGTGTTCAGGGCCAGCCCTGGCTGGCTGGAGGGTGAGGCTGGGTGT
1861    ------------------------------------------------------------  1920
        AACCTAGGGTCCCCTCACAAGTCCCGGTCGGGACCGACCGACCTCCCACTCCGACCCACA

         G  S  Q  G  S  V  Q  G  Q  P  W  L  A  G  G  *  G  W  V  S -

                    N            N                    BE P     SSS
                    l            l              ADNscMpMANPaacSS
                    a            a              vrltonunvlsuurii
                    3            3              aaaNRlMlaas99Fnn
                                               2241211124166111

                                               /  /   //////
        CTGGAAGCATGGAGCATGGGACTGTTCTTTTACAAGACAGGACCCTGGGACCACAGAGGG
1921    ------------------------------------------------------------  1980
        GACCTTCGTACCTCGTACCCTGACAAGAAAATGTTCTGTCCTGGGACCCTGGTGTCTCCC

         G  S  M  E  H  G  T  V  L  L  Q  D  R  T  L  G  P  Q  R  A -

                                    S      S              S
                                    f      f          M M  MNDFaXFF  F
                                    a      a          n n  bdpouhoo  o
                                    N      N          l l  oenk3okk  k
                                    1      1          1 1  1211A211  1
                                                       /    //
        CAGGAACTTGCACAAAATCACACAGCCAAGCCAGTCAAGGATGGATGCAGATCCAGAGGT
1981    ------------------------------------------------------------  2040
        GTCCTTGAACGTGTTTTAGTGTGTCGGTTCGGTCAGTTCCTACCTACGTCTAGGTCTCCA

         G  T  C  T  K  S  H  S  Q  A  S  Q  G  W  M  Q  I  Q  R  F -

            F                                 P               AFNF
            n                R B BB B     M M HnN H           vol
            u                s b bb b     n n pul h           akak
            4                a v vv v     l l h4a b           2141
            H                1 1 11 1     1 1 1H3 1            //
                                          /
        TTCTGGCAGCCAGTACCTCCTGCCCCATGCTGCCCGCTTCTCACCCTATGTGGGTGGGAC
2041    ------------------------------------------------------------  2100
        AAGACCGTCGGTCATGGAGGACGGGGTACGACGGGCGAAGAGTGGGATACACCCACCCTG

         L  A  A  S  T  S  C  P  M  L  P  A  S  H  P  M  W  V  G  P -

            S    H                                   M M       NR
            aS   i                                   n n       ls
            ui   n                                   l l       aa
            9n   f                                   1 1       31
            61   1
            //
        CACAGACTCACATCCTGACCTTGCACAAACAGCCCCTCTGGACACAGCCCCATGTACACG
2101    ------------------------------------------------------------  2160
        GTGTCTGAGTGTAGGACTGGAACGTGTTTGTCGGGGAGACCTGTGTCGGGGTACATGTGC

         Q  T  H  I  L  T  L  H  K  Q  P  L  W  T  Q  P  H  V  H  G -
```

— — — FIG. 3(— — — —)

9/93

FIG. 3(cont'd)

```
 H      F    P    M  M              P    P         M  M         D
 a      o    o    n  n              o    o         n  n         d
 e      k    k    l  l              k    k         l  l         e
 3      1    1    1  1              1    1         1  1         1
       GCCTCAAGGGATGTCTCACATCCTCTGTCTATTTGAGACTTAGAAAAATCCTACAAGGCT
2161   --------------------------------------------+---------+--------  2220
       CGGAGTTCCCTACAGAGTGTAGGAGACAGATAAACTCTGAATCTTTTTAGGATGTTCCGA

         L   K   G   C   L   T   S   S   V   Y   L   R   L   R   K   I   L   Q   G   W -
```

```
                                       S                                   B H
                                    BMDNa                          M M  ADBsgNSS
           A              D         cbpdu                          n n  ldapisas
           c              d         lone3                          l l   uenlApct
           c              e         1112A                          l l  11221211
           1              1          / //                                 /////
       GGCAGTAGACAGAACTAAGATGATCATCTCCAGTTTATAGACCAGAACCAGAGCTCAGAG
2221   --------------------------------------------+---------+--------+  2280
       CCGTCATCTGTCTTGATTCTACTAGTAGAGGTCAAATATCTGGTCTTGGTCTCGAGTCTC

         Q   •   T   E   L   R   •   S   S   P   V   Y   R   P   E   P   E   L   R   E -
```

```
                                                              H            BES
                                                              i            scc
                           HM    M                            n            tor
           M               ps    a                            f            NRP
           a               ap    e                            1            121
           e               21    1                                           /
           1                /
       AGGCTAGATGATTGATTACCAAGTGCCGGACTAGCAAGTGCTGGAGTCGGGACTAACCCA
2281   --------------------------------------------+---------+--------+  2340
       TCCGATCTACTAACTAATGGTTCACGGCCTGATCGTTCACGACCTCAGCCCTGATTGGGT

         A   R   •   L   I   T   K   C   R   T   S   K   C   W   S   R   D   •   P   R -
```

```
         P S                        P              B M MB          H      H
       ADNpPaS        B    B        nu             s n ns          P      P
       vrlusui        b    b        u4             m l lm          h      h
       aaaMs9n        V    V        H              1 1 11          1      1
       2241161        1    1
        / ////
       GGTCCCTTGTCCCAAGTTCCACTGCTGCCTCTTGAATGCAGGGACAAATGCCACACGGCT
2341   --------------------------------------------+---------+--------+  2400
       CCAGGGAACAGGGTTCAAGGTGACGACGGAGAACTTACGTCCCTGTTTACGGTGTGCCGA

         S   L   V   P   S   S   T   A   A   S   •   M   Q   G   Q   M   P   H   G   S -
```

```
                       M                 R                 R
                       a                 s                 s
                       e                 a                 a
                       1                 1                 1
       CTCACCAGTGGCTAGTGGTGGGTACTCAATGTGTACTTTTGGGTTCACAGAAGCACAGCA
2401   --------------------------------------------+---------+--------+  2460
       GAGTGGTCACCGATCACCACCCATGAGTTACACATGAAAACCCAAGTGTCTTCGTGTCGT

         H   Q   W   L   V   V   G   T   Q   C   V   L   L   G   S   Q   K   H   S   T -
```

```
                           S                               HS      P    P  MFM
         SN    N         ANaS         D                     at      n    n  lkl
         tc    l         vlui         d                     eu      k    k  111
         Y     a         aa9n         e                     31      1    1  111
         11    3         2461         1                      /
       CCCATGGGAAGGGTCCATCTCAGAGAATTTACGAGCAGGGATGAAGGCCTCCCTGTCTAA
2461   --------------------------------------------+---------+--------+  2520
       GGGTACCCTTCCCAGGTAGAGTCTCTTAAATGCTCGTCCCTACTTCCGGAGGGACAGATT

         H   G   K   G   P   S   Q   R   I   Y   E   Q   G   •   R   P   P   C   L   K -
```

SUBSTITUTE SHEET

10/93

FIG. 3(cont'd)

```
      F                 M  M         N              H        M  MM                      M         M
      o                 n  n         s              i        n  n  na                   b         b
      k                 l  l         P              n        n  l  le                   o         o
      1                 1  1         2              f        1     13                    2         2
     AATCCCTCCTTCATCCCCCGCTGGTGGCAGAATCTGTTACCAGAGGACAAAGCCTTTGGC
2521 ------------------------------------------------------------ 2580
     TTAGGGAGGAAGTAGGGGGCGACCACCGTCTTAGACAATGGTCTCCTGTTTCGGAAACCG

         S    L    L    H    P    P    L    V    A    E    S    V    T    R    G    Q    S    L    W    L -


                   H        H      A            BH              P          B          B   M
                   i        h      l            sgN             s          s          s   a
                   n        a      u            pia             t          m          m   e
                   P        1      1            1Ap             1          1          1   3
                   1                            212                                     
                                                 //
     TCTTCTAATCAGAGCGCAAGCTGGGAGCACAGGCACTGCAGGAGAGAATGCCCAGTGACC
2581 ------------------------------------------------------------ 2640
     AGAAGATTAGTCTCGCGTTCGACCCTCGTGTCCGTGACGTCCTCTCTTACGGGTCACTGG

         F    *    S    E    R    K    L    G    A    Q    A    L    Q    E    R    M    P    S    D    Q -


                              BES                                                            MAMMN
                              scc         M  M     AM  M                                      nlanb
       M                      tor         n  n     ln  n                                      luele
       a                      NRP         l  l     ul  l                                      11111
       e                      121         1  1     11  1                                        //
       3                        /
     AGTCACTGACCCTGTGCAGAACCTCCTGGAAGCGAGCTTTGCTGGGAGAGGGGGTAGCTA
2641 ------------------------------------------------------------ 2700
     TCAGTGACTGGGACACGTCTTGGAGGACCTTCGCTCGAAACGACCCTCTCCCCCATCGAT

         S    L    T    L    C    R    T    S    W    K    R    A    L    L    G    E    G    V    A    S -


                                       P  S                      BES               H
                                   ADMMNpPaS       M  M          HsccH             i     H
         D              N      D   vrnnlusui       n  n          ptorp             n     h
         d              l      d   aallaMs9n       l  l          hNRPh             P     a
         e              a      e   221141161       1  1          11211             1     1
         1              4      1      // //////                    /
     GCCTGAGAGGGAACCCTCTAAGGGACCTCAAAGGTGATTGTGCCAGGCTCTGCGCCTGCC
2701 ------------------------------------------------------------ 2760
     CGGACTCTCCCTTGGGAGATTCCCTGGAGTTTCCACTAACACGGTCCGAGACGCGGACGG

         L    R    G    N    P    L    R    D    L    K    G    D    C    A    R    L    C    A    C    P -


                              M  M       M  MM  M                           M      HM
                              n  n       n  nn  n                           b      ab
                              l  l       l  11  l                           o      eo
                              1  1       1  11  1                           2      32
     CCACACCCTCCCTTACCCTCCTCCAGACCATTCAGGACACAGGGAAATCAGGGTTACAAA
2761 ------------------------------------------------------------ 2820
     GGTGTGGGAGGGAATGGGAGGAGGTCTGGTAAGTCCTGTGTCCCTTTAGTCCCAATGTTT

         T    P    S    L    T    L    L    Q    T    I    Q    D    T    G    K    S    G    L    Q    I -


                    S                 B                   S
                  MNDa        D  M   aMDMNNaX     H  M   HM            M  M  D
                  bdpu        d  b   mbpbdluh     P  n   pn            n  n  d
                  en3         e  H   noea3        h  1   h1            l  l  e
                  121A        1  2   111224A2     1  1   11            1  1  1
                   / /              /  //////                 /
     TCTTCTTGATCCACTTCTCTCAGGATCCCCTCTCTTCCTACCCTTCCTCACCACTTCCCT
2821 ------------------------------------------------------------ 2880
     AGAAGAACTAGGTGAAGAGAGTCCTAGGGGAGAGAAGGATGGGAAGGAGTGGTGAAGGGA

         F    L    I    H    F    S    Q    D    P    L    S    S    Y    P    S    S    P    L    P    S -
```

11/93

FIG.3(cont'd)

```
                                                                    BES
      M M                                        MMADMM             acc
      c D                                        nbhrnb             tor
      l l                                        loaalo             NRF
      1 1                                        123112             121
                                                   //                 /
      CAGTCCCAACTCCTTTTCCCTATTTCCTTCTCCTCCTGTCTTTAAAGCCTGCCTCTTCCA
2881  ------------+---------+---------+---------+---------+---------+   2940
      GTCAGGGTTGAGGAAAAGGGATAAAGGAAGAGGAGGACAGAAATTTCGGACGGAGAAGGT

       V  P  T  P  F  P  P  Y  F  L  L  L  L  S  L  R  P  A  S  S  R  -
```

```
                           F                   B                      B
      M  MB   B   M    M    n          NBsN                            sN
      n  nb   b   b    b    u          laps                            ps
      l  lv   v   o    o    4          anlp                            lp
      1  11   1   2    2    H          4222                            22
                                         //                            /
      GGAAGACCCCCCTATTGCTGCTGGGGCTCCCCATTTGCTTACTTTGCATTTGTGCCCACT
2941  ------------+---------+---------+---------+---------+---------+   3000
      CCTTCTGGGGGGATAACGACGACCCCGAGGGGTAAACGAATGAAACGTAAACACGGGTGA

       R  T  P  L  L  L  L  G  L  P  I  C  L  L  C  I  C  A  H  S  -
```

```
                        D        A
                        d        l
                        e        u
                        l        l
      CTCCACCCCTGCTCCCCTGAGCTGAAATAAAAATACAATAAACTTACTATAAAGATGAAA
3001  ------------+---------+---------+---------+---------+---------+   3060
      GAGGTGGGGACGAGGGGACTCGACTTTATTTTTATGTTATTTGAATGATATTTCTACTTT

       P  P  L  L  P  *  A  E  I  K  I  Q  *  T  Y  Y  R  D  E  K  -
```

```
          AAAA
3061      ----   3064
          TTTT

          ?  -
```

Enzymes that do cut:

Accl	Aha2	Aha3	Afl2	Alul	Aval	Ava2	BamH1
Ban1	Ban2	Bbel	Bbv1	Bcl1	Bsml	Bsp12	BspM1
BstE2	BstN1	BstX1	Ddel	Dpnl	Dra1	Dra2	EcoB
EcoR2	Espl	PnuD2	Fnu4H	Fokl	Hae2	Hae3	Hga1
HgiA1	HgiD1	Hha1	Hinf1	HinP1	Hpa2	Hph1	Mael
Mae3	Mbol	Mbo2	Mnll	Mspl	Mst2	Nael	Narl
Ncil	Ncol	Nde2	Nhel	Nla3	Nla4	Nsp2	NspB2
NspH1	PflM1	PpuM1	Pssl	Pst1	Pvu2	Rsal	Sacl
Sau3A	Sau96	ScrF1	SfaN1	Sinl	Sst1	Stul	Styl.
Taq1	Thal	Xho2					

Enzymes that do not cut:

Aat2	Apal	ApaL1	Asp70	Asp71	Asu2	Avr2	Ball
Bgl1	Bgl2	BspM2	BssH2	Cfrl	Clal	Dra3	Eagl
EcoK	EcoK	EcoR1	EcoRV	Pspl	Hinc2	Hind3	Hpal
Kpnl	Mae2	Mlul	Mstl	Ndel	Notl	Nrul	Nsil
PaeR7	Pvul	Rsr2	Sac2	Sall	Scal	Sfil	Smal
SnaBl	Spel	Sphl	Sspl	Sst2	Tthl	Xbal	Xhl
Xmal	Xma3	Xmnl	Xor2				

12/93   *FIG. 4*

T4 cDNA
EcoRI |PstI (~3kbp)   PstI| EcoRI

λgt 10.PBL.T4
= λ203-4

↓ *EcoRI, isolate 3 kbp fragment*

↓── pBG 312· *EcoRI* (4.75 kbp)

↓ *ligate w/*
T4 ligase

T4
PstI
EcoRI   PstI
                                                   EcoRI
PstI

Amp^r

pBG 312.T4
(p 171-1)
7.75 kbp

ori
pA

↓ *PstI, klenow, isolate 25 kbp fragment*

↓── pBG 312· *SmaI* (4.75 kbp)

ligate w/T4 ligase

EcoRI

E SV40   AdMLP   EcoRI
                      (SmaI/PstI)

Amp^r

p170-2
7.4 kbp

T4

Ava I

(SmaI/PstI)
ori
pA   BglII

FIG. 5

1(19)* 104/394 (123/423)   1,200 (1,604)

splice
error   T4

EcoRI

Lac Z

ori

pEC 100
3.9 Kbp

Amp^r
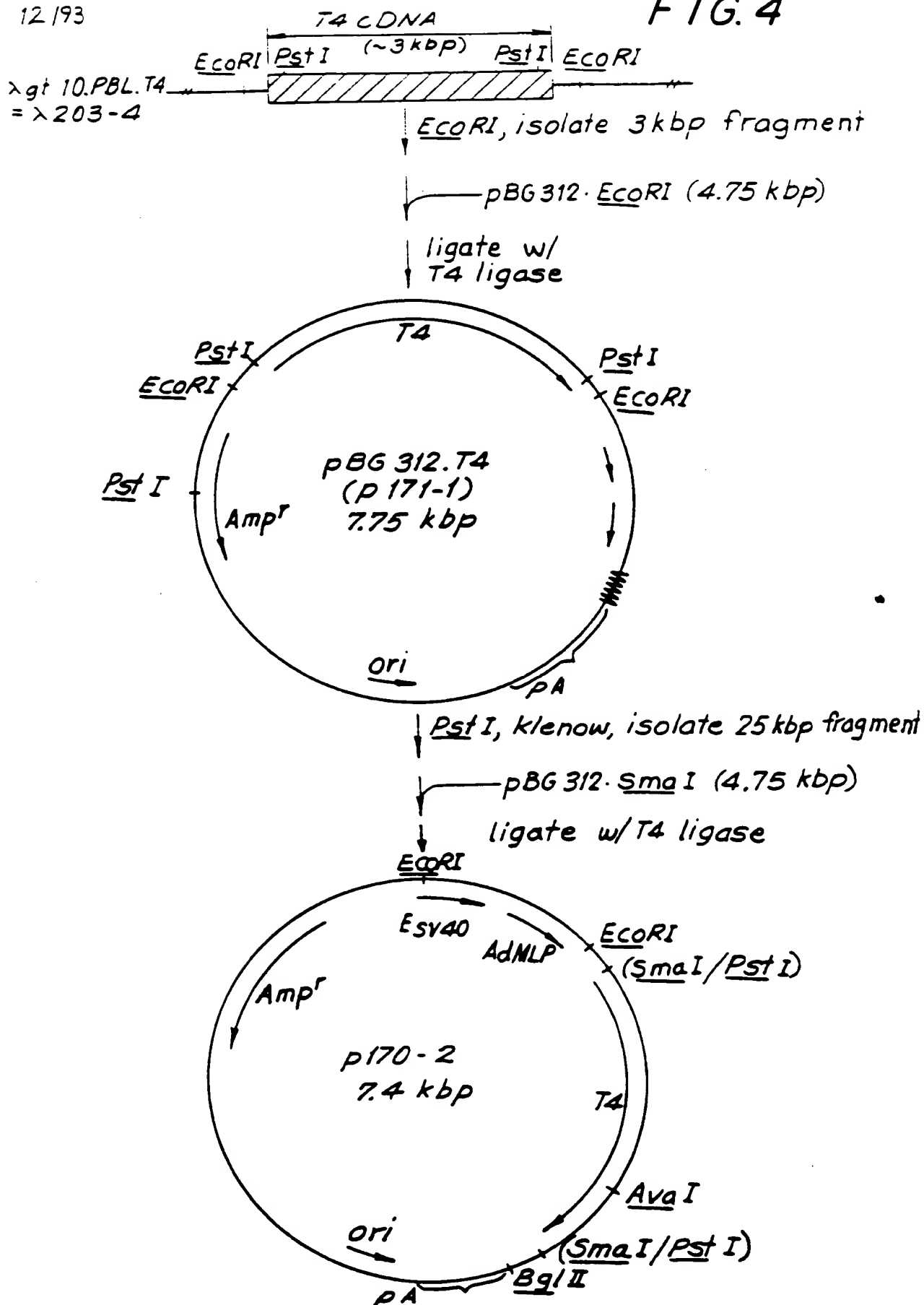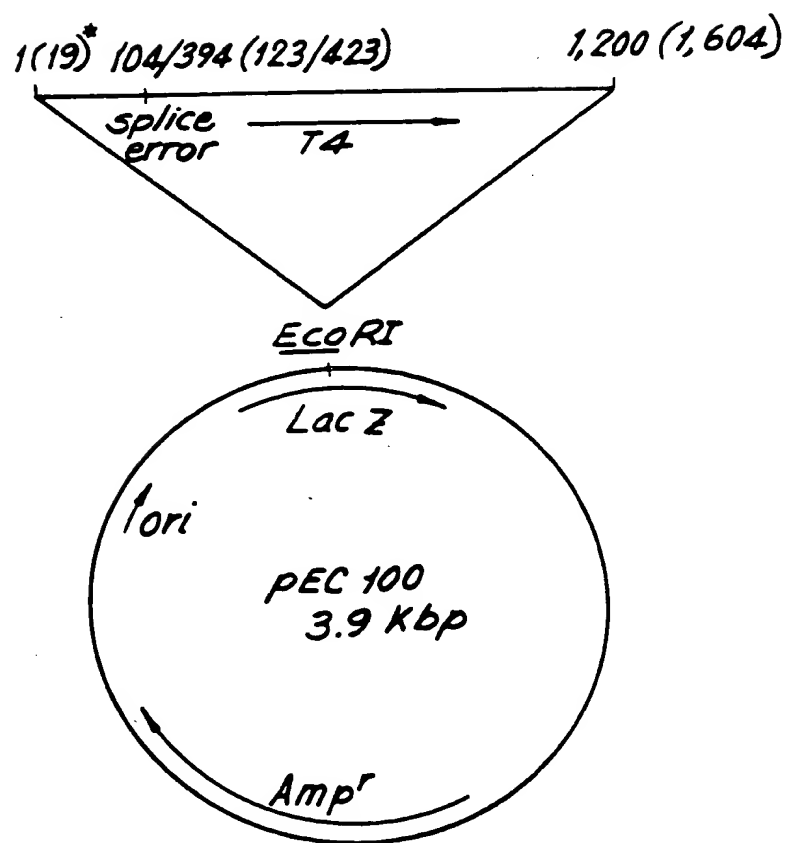
* numbers in parentheses refer to PBL T4 cDNA coordinates

14/93

FIG.6

AMINO ACID SEQUENCE COMPARISON AT POSITIONS 3,6 AND 231 OF T4

Position No.	Maddon et al	PBL Clone	Rex Clone	Genomic	Mouse	Sheep
3	N AAC	K AAG	–	–	K AAG	K –
64	W TGG	R CGG	W TGG	W TGG	W TGG	–
231	F TTT	S TCT	F TTT	–	F TTC	–

FIG. 7A

FIG. 7B

17/93
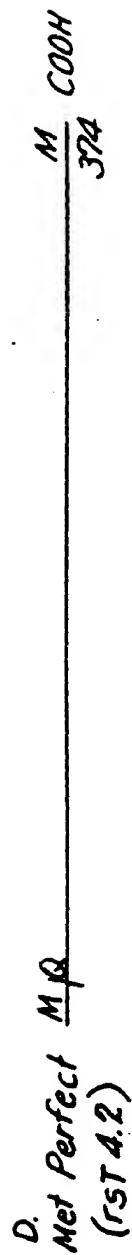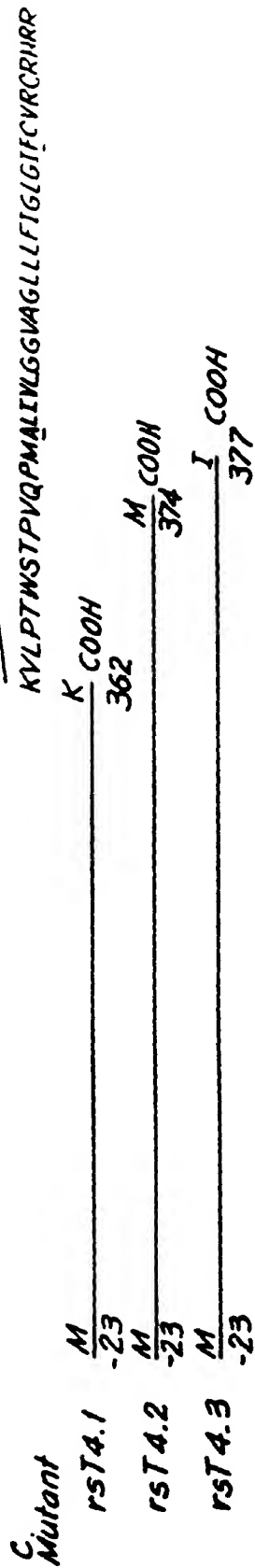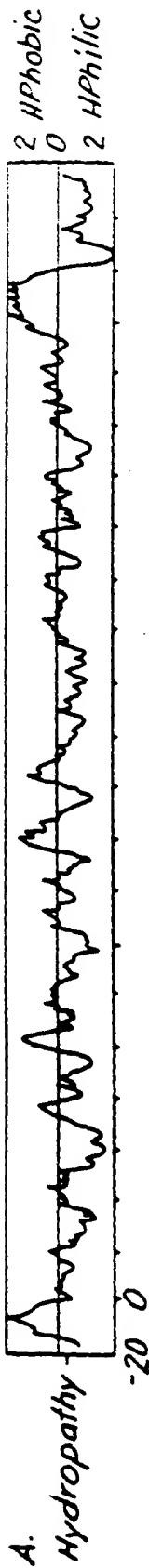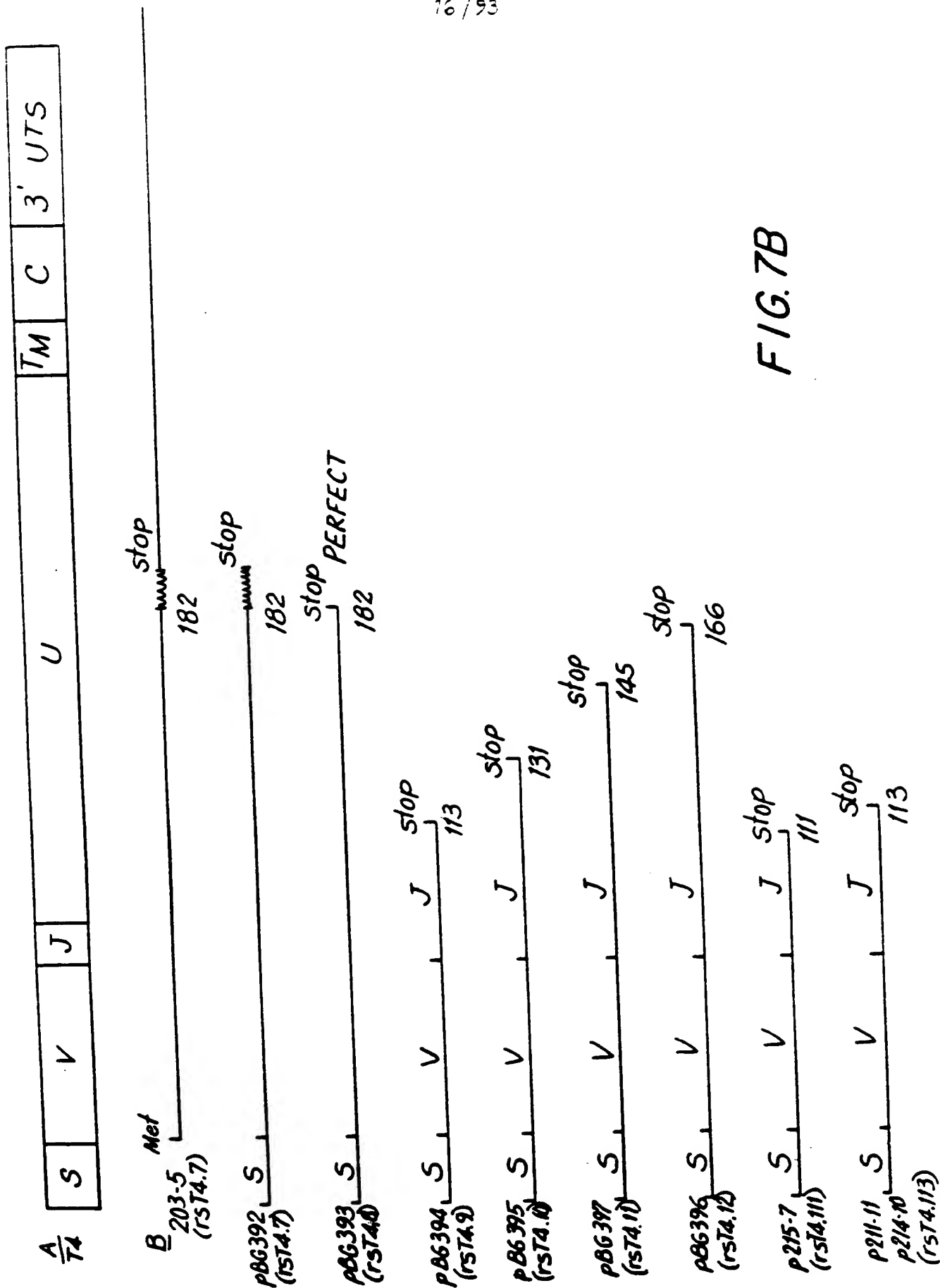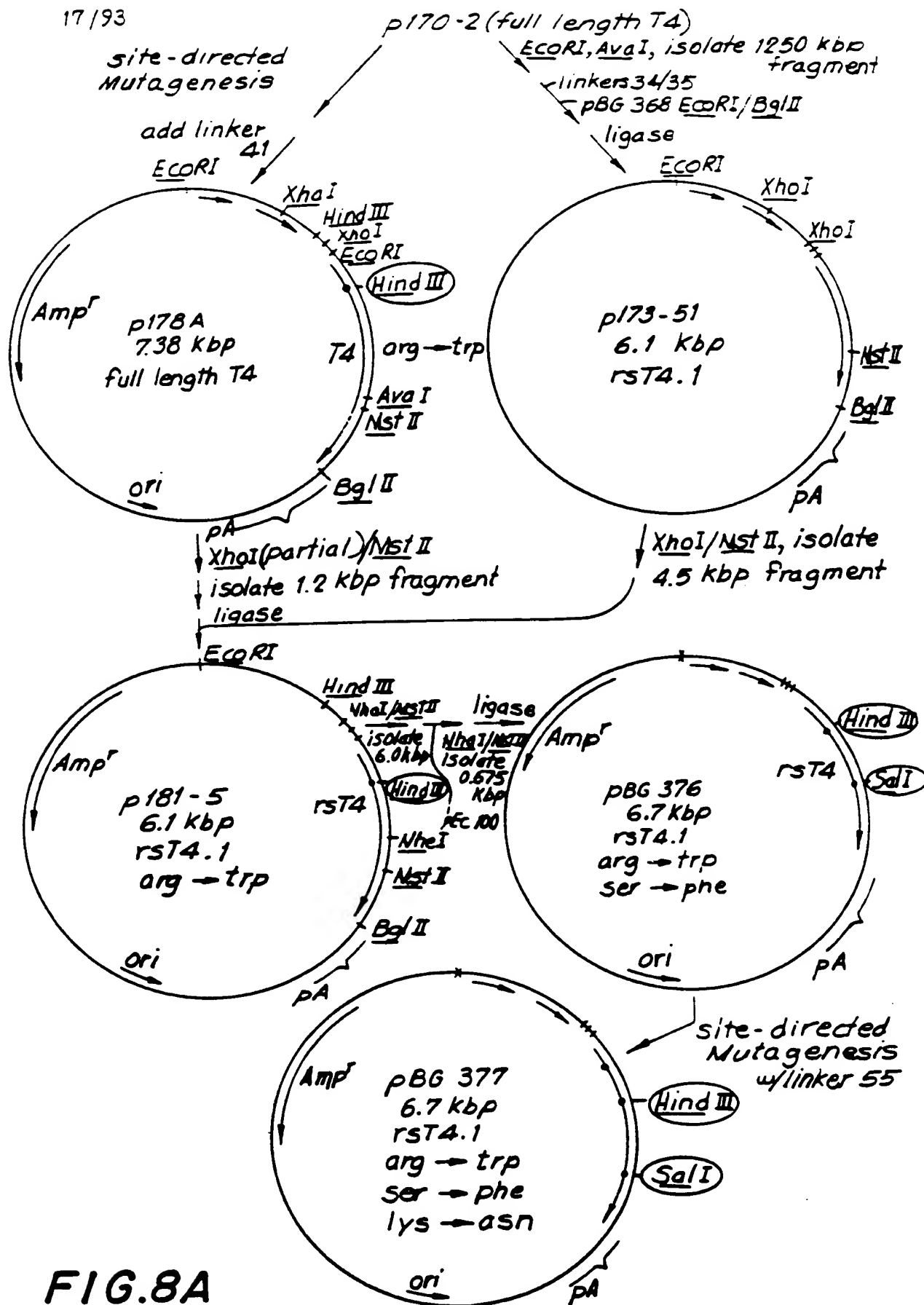
FIG.8A

FIG.8B   p178A (full length T4)

arg → trp |Ava I, EcoRI, Isolate 1.2 Kbp fragment

⊢— Linkers 46/47, 48/49
    ⟍ pBG 368 · EcoRI/Bgl II

| Ligase

(EcoRI)

p185-23
6.14 Kbp
rsT4.2
(arg → trp)

Amp^r

ori

EcoRI

Hind III

rsT4

Ava I

Sal I

Bgl II

PA
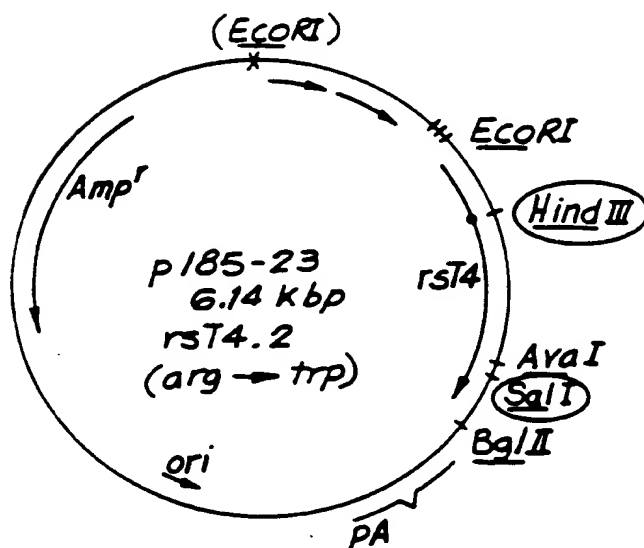
| EcoRI/SalI, Isolate 1.3 kbp fragment

⊢— Linkers 50/51
    ⟍ pBG 368 · EcoRI/Bgl II

p186-6
6.15 Kbp
rsT4.3
arg → trp

Amp^r

ori

Hind III

rsT4

Sal I

Bgl II

PA
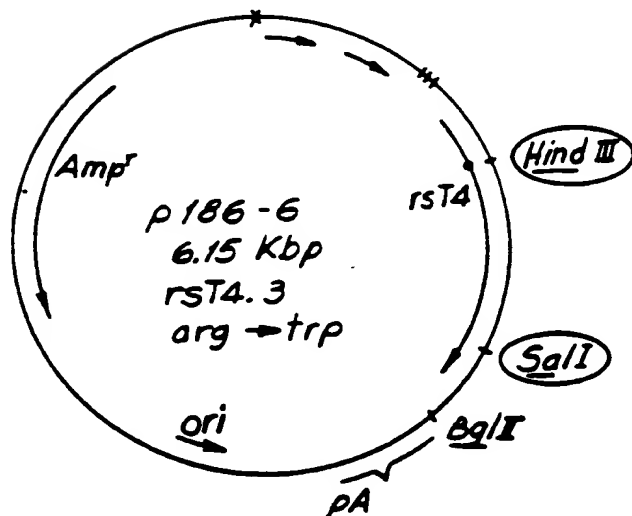
"Met-perfect" cassette for E.coli expression of rsT4
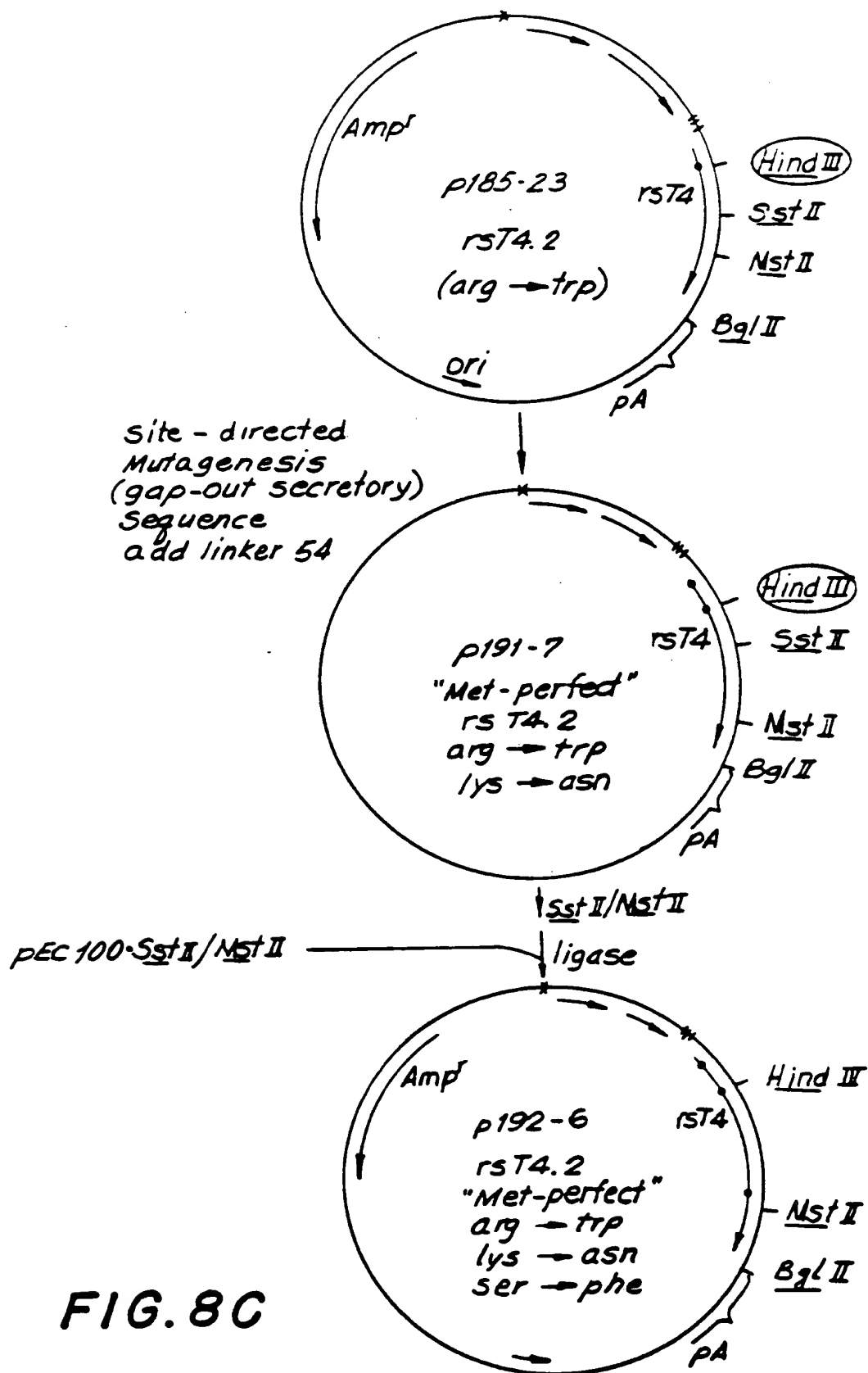
FIG. 8C

FIG.8D

p1034 · *Bam*HI / *Bgl*II
_____
isolate
terminator
fragment

192·6·*Bgl*II

ligase

p193-8
7.24 Kbp

Amp^r

*Cla*I
*Bal*I
rsT4

φT4
term

*Bgl*II/*Bam*HI

TcR

*Cla*I

*Bgl*II/*Bgl*II

*Aha*III/*Eco*RI

Promoter
*Cla*I

Signal

rop⁻ ST8
(rsT4)

rsT4

TcR

φT4
term

*Pvu*II
*Bal*I   *Ava*I   *Sal*I   *Cla*I
            *Hind*III

p194-3
4.03 kbp

φT4
Promoter

*Cla*I
*Bal*I
*Xho*I

rsT4

φT4
term

TcR

*Cla*I

p194-3   *Bal*I + *Xho*I
Mung Bean Nuclease
to blunt *Xho*I

ligase

p195-8
3.94 Kbp
E.coli
expression
construct

φT4
Promoter

*Cla*I

rsT4

φT4
term

TcR

*Cla*I
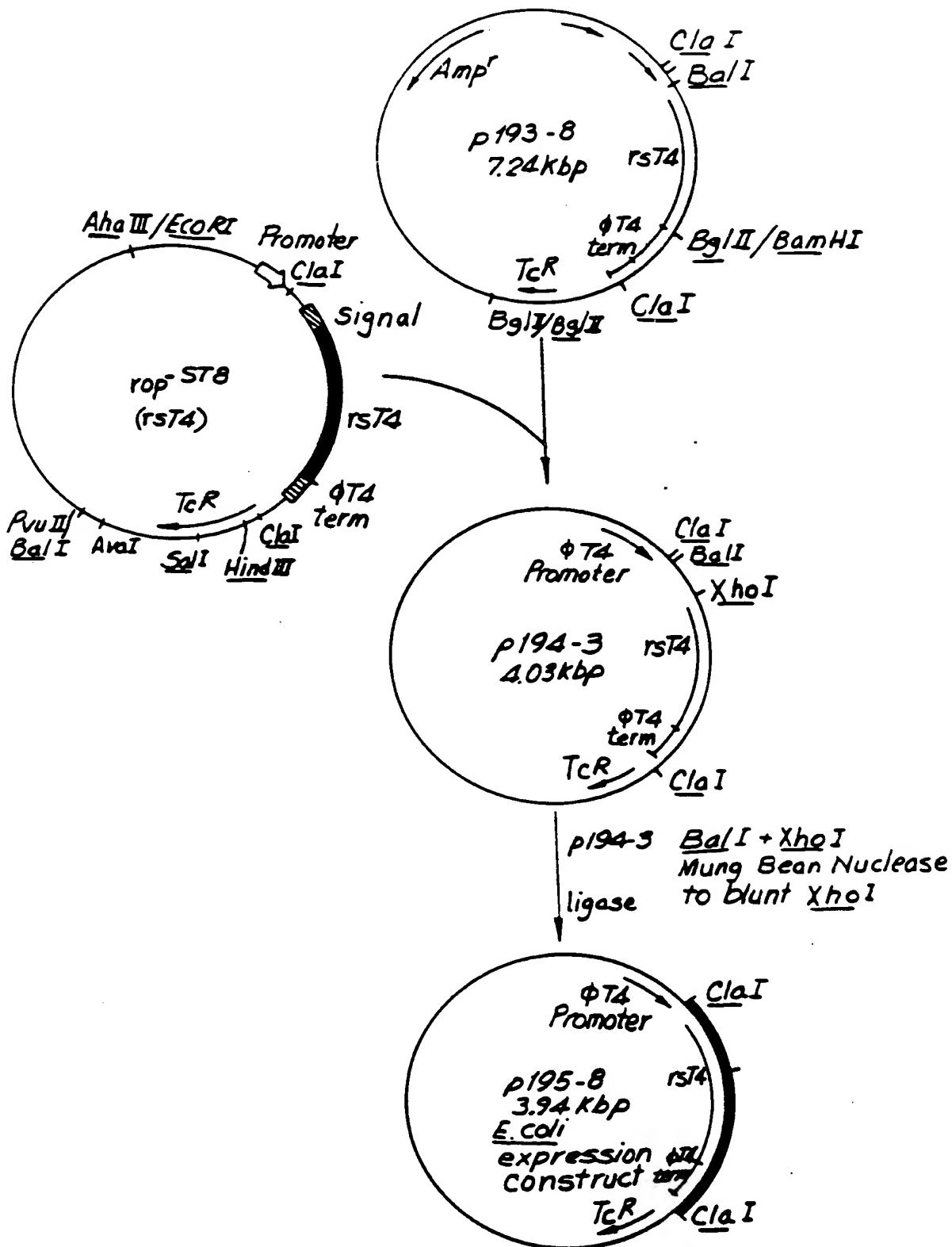
21/93

FIG. 9A
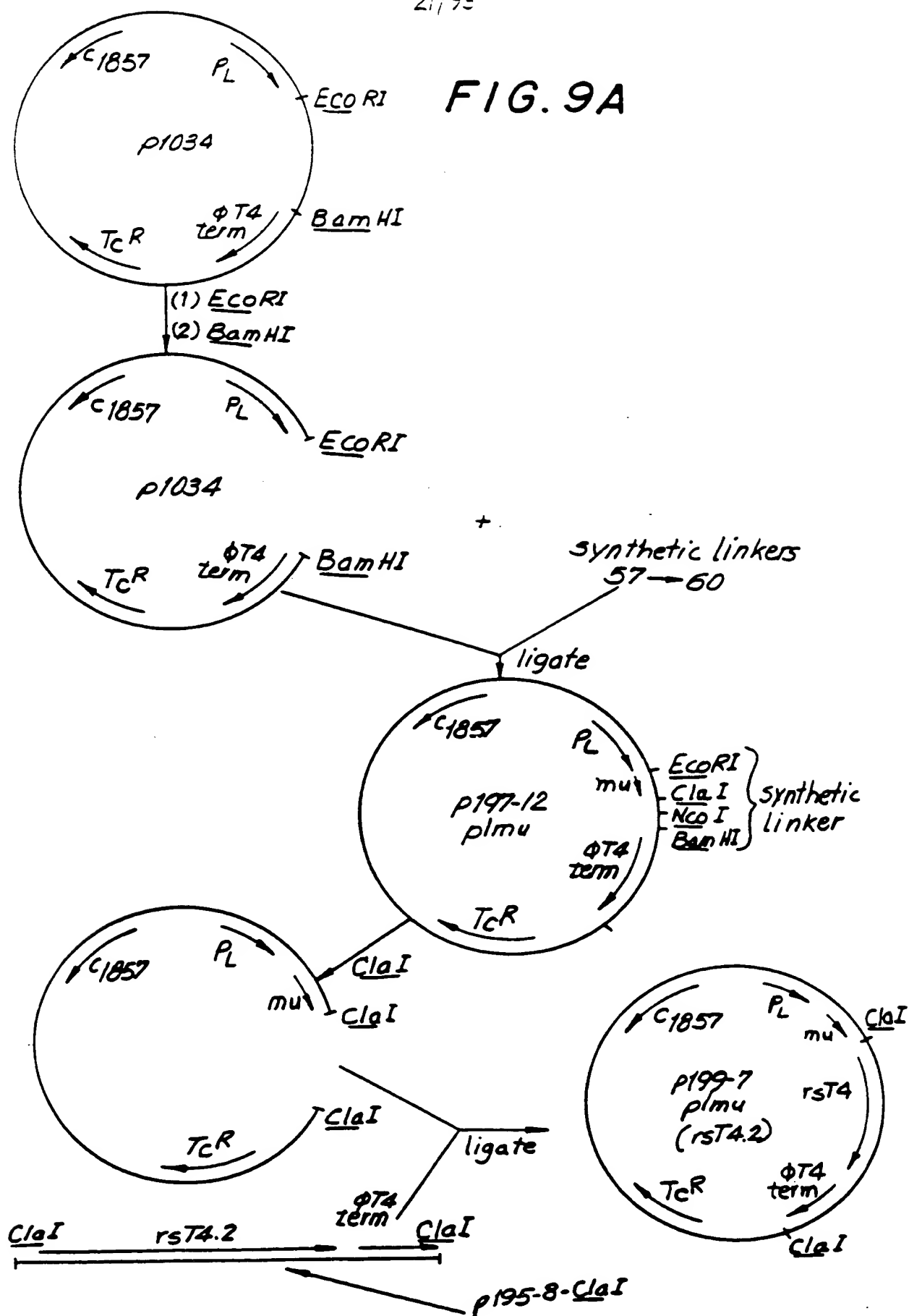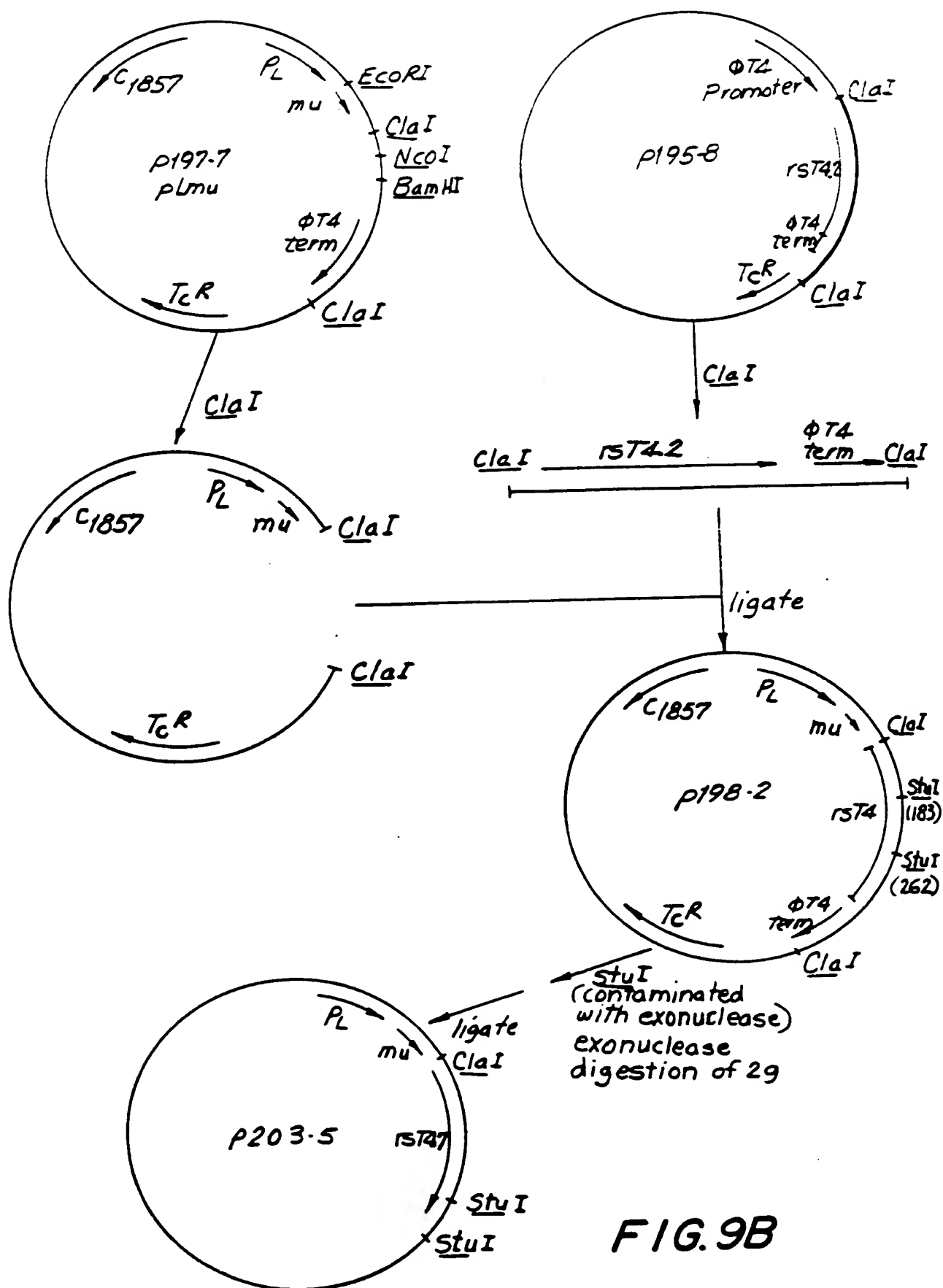
FIG. 9B

FIG. 9C

198-2

```
          180                      StuI
    CTA   GCT   TTC   CAG   AAG    GCC   TCC   AGC   ATA   GTC   TAT
```

deletion

203-5

```
          180                182
    CTA   GCT   TTC   CAG   AAC   CTC   CAG   CAT   AGT   CTA   TAA
    leu   ala   phe   gln   asn   leu   gln   his   ser   leu   stop
```

←— T4 frameshift —→

24/93

FIG. 10

<u>48</u>

5' CTG CCC ACA TGG TCC ACC CCG GTG CAG CCA ATG TGA 3'

<u>49</u>

5' CAT CTC ACA TTG GCT GCA CCG GGG TCG ACC ATG T 3'

<u>50</u>

5' TCG ACC CCG GTG CAG CCA ATG GCC CTG ATT TGA 3'

    3' GG GGC CAC GTC GGT TAC CGG GAC TAA ACT CTA G 5'

<u>51</u>

<u>41</u>

5' GAA GAA GCT TGT GGG ACC AAG 3'

<u>34</u>

5' TCG GGA CAG GTC CTG CTG GAA TCC AAC ATC AAG TGA A 3'

    3' CTG TCC AGG ACG ACC TTA GGT TGT AGT TCA CTT CTA G 5'

<u>35</u>

<u>55</u>

5' CAG CCA CCC AAG GAA ACA AAG TGG 3'

<u>46</u>

5' TCG GGA CAG GTC CTG CTG GAA TCC AAC ATC AAG GTT 3'

<u>47</u>

5' GGG CAG AAC CTT GAT GTT GGA TTC CAG CAG CAC CTG TC 3'

FIG. 10(cont'd)

54

5' AGC TTC GAC TCG AGG ATG CAG GGA AAC AAA CTC GTG 3'

57   **58**

5' AATTCTTACACTTACTTAAATTGCTAACTTTATAGATTACAAAACTT

GAATGTGAATCAATTTAACGATTGAAATATCTAATGTTTGAA
                                              59

AGGAAATCGATTTCCATGG

TCCTTTAGCTAAAGGTTACCCTAG 5'
                60

FIG.11

```
E                    C                              D        ClaI
c                    d                              d
o                    l                              e
R                    l                              I    RBS
I                                                   I
GAATTCTTACACTTAGTAAATTGCTAATTTATAGATACAAACTTAGGAAATCATC     60
1  ----------------+---------+---------+---------+---------+
   CTTAAGAATGTGAATCATTTAACGATTAAATATCTATGTTTGAATCCTTTAGCTAG

   GluPheLeuHisLeuValLysLeuLeuIleLeuAsnIleThrLysLeuArgLysSerIle -
```

```
       F  E                                        R
       o  c                                        s
       k  o                                        a
       I  R                                        I
TGGATGCCAGGGAAACAAAGTGGTCCTGGGCAAAAAGGGGATACAGTGGAACTGACCTGT    120
61 ----------------+---------+---------+---------+---------+
   ACCTACGGTCCCTTTGTTTCACCAGGACCCGTTTTTCCCCTATGTCACCTTGACTGGACA

   TrpMetGlnGlyAsnLysValValLeuGlyLysLysGlyAspThrValGluLeuThrCys -

       Initiating
       A  MET
       l
       u
       I
```

```
              M   M                                I
              b   b                                s
              o   o                                a
              2   2                                I
ACAGCTTCCCAGAAGAAGAGCATACAATTCCACTGGAAAAACTCCAACCAGATAAGATT    180
121 ----------------+---------+---------+---------+---------+
   TGTCGAAGGGTCTTCTTCTCGTATGTTAAGGTGACCTTTTTGAGGTTGGTCTATTCTAA

   ThrAlaSerGlnLysLysSerIleGlnPheHisTrpLysAsnSerAsnGlnIleLysIle -
```

```
   S                  S            A       S  FN       I
MfaM  F              AaS           l     MfoDa aInT      s
lapa  o              vuI           s     bdpu  unhh      a
asIp  k              a9E           I     oeaJ  DPaa      I
4223  I              361          1     121A  2111     1
CTGGGAAATCAGGGCCTCCTTCTTAACTAAAGGTCCATCCAGCTGAATGATCGGCCTGAC    240
181 ----------------+---------+---------+---------+---------+
   GACCCTTTAGTCCCGGAGGAAGAATTGATTTCCAGGTAGGTCGACTTACTAGCCGGACTG

   LeuGlyAsnGlnGlySerPheLeuThrLysGlyProSerLysLeuAsnAspArgAlaAsp -
```

```
  H        S            S          E     A
  i    A  MfaMaSS        BXDNa      s     l
  n    l  bvluIt         cbpdu      a     f
  d    u  oaaFRY         loaeJ      I     I
  3    I  234611        1112A      I     I
TCAGAAGAACCTTGTGGGCACCAAGGAAACTTCCCCTGATCATCAAGAATCTTAAGATA    300
241 ----------------+---------+---------+---------+---------+
   AGTCTTCTTGGAACACCCTGGTTCCTTTGAAGGGGACTAGTAGTTCTTAGAATTCTAT

   SerArgArgSerLeuTrpAspGlnGlyAsnPheProLeuIleIleLysAsnLeuLysIle -
```

```
  H     D          M       M      S            M
  i     d          b       s    MfaMS          a
  n                o       a    avuai          e
  d                2       I    la9Ia          I
  3     I                  I    12611
GAAGACTCAGATACTTACATCTGTGAAGTGGAGGACCAGAAGGAGGAGGTGCAATTGCTA    360
301 ----------------+---------+---------+---------+---------+
   CTTCTGAGTCTATGAATGTAGACACTTCACCTCCTGGTCTTCCTCCTCCACGTTAACGAT

   GluAspSerAspThrTyrIleCysGluValGluAspGlnLysGluGluValGlnLeuLeu -
```

FIG.11(cont'd)

```
                                   E              B
                                   c              s
                                   o              p
                                   B              M
                                                  1
         GTGTTCGGATTGACTGCCAACTCTGACACCCACCTCCTTCAGGGGCAGAGCCTTACCCTG
361      ------------------------------------------------------------ 420
         CACAAGCCTAACTGACGGTTGAGACTGTGGGTGGACGAAGTCCCCGTCTCGGACTGGGAC
         ValPheGlyLeuThrAlaAsnSerAspThrHisLeuLeuGlnGlyGlnSerLeuThrLeu -


                 BBE S                              H
                 BsscNc                D        M    i       S
         s       aptosr                d        n    n       c
         t       nlNRpF                d        l    f       Y
         Y       221221                l        l    1       1
                  / ///
         ACCTTGGAGAGCCCCCCTGGTAGTAGCCCCTCACTGCAATGTAGGAGTCCAAGGGGTAAA
421      ------------------------------------------------------------ 480
         TGGAACCTCTCGGGGGGGACCATCATCGGGGAGTCACGTTACATCCTCAGGTTCCCCATTT
         ThrLeuGluSerProProGlySerSerProSerValGlnCysArgSerProArgGlyLys -


                                       N        A BBEH SB      BES
                            M  M  D    AsP      A BsscGNScsS B Nscc
                            b  n  d    lpv      l aptoisarts a ltor
                            o  l  e    uBu      u nlNRApcFXt n aNRF
                            2  1  1    122      1 2212121111 1 4121
                                        //       /////// /        /
         AACATACAGGGGGGGAAGACCCTCTCCGTGTCTCAGCTGGAGCTCCAGGATAGTGGCACC
481      ------------------------------------------------------------ 540
         TTGTATGTCCCCCCCTTCTGGGAGAGGCACAGAGTCGACCTCGAGGTCCTATCACCGTGG
         AsnIleGlnGlyGlyLysThrLeuSerValSerGlnLeuGluLeuGlnAspSerGlyThr -


                 N                        M                      N
                 Ns                       b                      h
                 lp                       o                      e
                 aH                       2                      1
                 31
                  /
         TGGACATGCACTGTCTTGCAGAACCAGAAGAAGGTGGAGTTCAAAATAGACATCGTCGTC
541      ------------------------------------------------------------ 600
         ACCTGTACGTGACAGAACGTCTTGGTCTTCTTCCACCTCAAGTTTTATCTGTAGCAGCAG
         TrpThrCysThrValLeuGlnAsnGlnLysLysValGluPheLysIleAspIleValVal -


         M  A              HS        M     M
         a  l              at        b     b
         e  u              eu        l     l
         1  1              31        1     1
                            /
         CTAGCTTTCCAGAAGGCCTCCAGCATAGTCTATAAGAAAGAGGGGGAACAGGTGGAGTTC
601      ------------------------------------------------------------ 660
         GATCGAAAGGTCTTCCGGAGGTCGTATCAGATATTCTTTCTCCCCCTTGTCCACCTCAAG
         LeuAlaPheGlnLysAlaSerSerIleValTyrLysLysGluGlyGluGlnValGluPhe -


                                       A                A      M
                                       l                l      b
                                       u                u      l
                                       1                1      1
         TCCTTCCCACTCGCCCTTTACAGTTGAAAAGCTGACGGGCAGTGGCGAGCTGTGGTGGCAG
661      ------------------------------------------------------------ 720
         AGGAAGGGTGAGCGGGAAATGTCAACTTTTCGACTGCCCGTCACCGCTCGACACCACCGTC
         SerPhePro LeuAlaPheThrValGluLysLeuThrGlySerGlyGluLeuTrpTrpGln -
```

FIG.II(cont'd)

```
                         P       S
                     MH  M:  MNDa                                      M
                     =p  al  bcpu                                      l
                     1h  1M  ce=3                                      o
                     11  11  121A                                      u
                      /     / /                                        N
     GCGGAGAGGGCCTTCCTCCTCAAGTCTTGCATCACCTTTGACCTGCAAGAACAAGGAAGTG
721  ------------------------------------------------------------  780
     CGCCTCTCCCGGAAGGAGGAGTTCAGAACGTAGTGGAAACTGGACTTCTTGTTCCTTCAC
     AlaGluArgAlaSerSerSerLysSerTrpIleThrPheAspLeuLysAsnLysGluVal -
```

```
         s     BES    P  S
         sM    scc  ADNpPDaS A                            A         H
         ta    tor  vrluedul l                            l         P
         Ke    NRF  aaaMse9B u                            u         b
         23    121  22411161 1                            .         1
          /     /    / //// //                            /          /
     TCTGTAAAACGGGTTACCCAGGACCCTAAGCTCCAGATGGGCAAGAAGCTCCCGCTCCAC
781  ------------------------------------------------------------  840
     AGACATTTTGCCCAATGGGTCCTGGGATTCGAGGTCTACCCGTTCTTCGAGGGCGAGGTG
     SerValLysArgValThrGlnAspProLysLeuGlnMetGlyLysLysLeuProLeuHis -
```

```
              BES                                        BES      S
         M    scc   HS      D        M          H        scc  MHa
         b    tor   at      d        b          P        tor  nau
         l    NRF   eu      e        l          b        NRF  le9
         1    121   31      1        1          1        121  136
          /    /    /                                     /    /
     CTCACCCTGCCCCAGGCCTTGCCTCAGTATGCTGGCTCTGGAAACCTCACCCTGGCCCTT
841  ------------------------------------------------------------  900
     GAGTGGGACGGGGTCCGGAACGGAGTCATACGACCGAGACCTTTGGAGTGGGACCGGGAA
     LeuThrLeuProGlnAlaLeuProGlnTyrAlaGlySerGlyAsnLeuThrLeuAlaLeu -
```

```
              S      BES                          H D   A
              f      scc                          P d   l
              a      tor                          b e   u
              M      NRF                          1 1   1
              1      121
                      /
     GAAGCCGAAAACAGGAAAGTTGCATCAGGAAGTGAACCTGGTGGTGATGAGAGCCACTCAG
901  ------------------------------------------------------------  960
     CTTCGGCTTTTGTCCTTTCAACGTAGTCCTTCACTTGGACCACCACTACTCTCGGTGAGTC
     GluAlaLysThrGlyLysLeuHisGlnGluValAsnLeuValValMetArgAlaThrGln -
```

```
                       P S   S
              M      ADMNpPaS f  D  MA       D
              b      vrlluuui a  d  al       d
              l      aaaaMs9B M  e  lu       e
              1      22441161 1  1  11       1
                     ///////
     CTCCAGAAAAATTTGACCTGTGAGGTGTGGGGACCCACCTCCCCTAAGCTGATGCTGAGT
961  ------------------------------------------------------------  1020
     GAGGTCTTTTTAAACTGGACACTCCACACCCCTGGGTGGAGGGGATTCGACTACGACTCA
     L uGlnLysAsnLeuThrCysGluValTrpGlyPr ThrSerProLysLeuMetLeuSer -
```

```
              M                      T                      M
              b                      a                      b
              l                      q                      l
              1                      1                      1
     TTGAAACTGGAGAACAAGGAGGCAAAGGTCTCGAAGCGGGAGAAGCCGGTGTGGGTGCTG
1021 ------------------------------------------------------------  1080
     AACTTTGACCTCTTGTTCCTCCGTTTCCAGAGCTTCGCCCTCTTCGGCCACACCCACGAC
     LeuLysLeuGluAsnLysGluAlaLysValSerLysArgGluLysAlaValTrpVaiL u -
```

FIG.11(cont'd)

```
                                              H                    P S          H
                                              i                    ADpPaS       i
          DM                    P    D    M    n    A              VTusui       n
          ds                    o    d    n    a    v              aaHs9n       f
          et                    k    e    a    e    a              221161       i
          12                    1    1    3    1    1              // //
          /
          AACCCTGAGGCGGGGATGTGGCAGTGTCTGCTGAGTGACTCGGGACAGGTCCTGCTGCAA
1081      ------------------------------------------------------------ 1140
          TTGGGACTCCGCCCCTACACCGTCACAGACGACTCACTGAGCCCTGTCCAGGACGACCTT
          AsnProGluAlaGlyMetTrpGlnCysLeuLeuSerAspSerGlyGlnValLeuLeuGlu -


                          H         S      P                 S        P
                          NSAiT     HHMiC  n                 BMDNaX    cP
                          lacna     psci   u                 bbpduh    ua
                          alccq     api7   4                 vone3o    4t
                          31121     3111   H                 1112A3    H1
                          /         ///                      /  ///
          TCCAACATCAAGGTTCTGCCCACATGGTCGACCCCGGTGCAGCCAATGTGAGATCCTGCA
1141      ------------------------------------------------------------ 1200
          AGGTTGTAGTTCCAAGACGGGTGTACCAGCTGGGGCCACGTCGGTTACACTCTAGGACGT
          SerAsnIleLysValLeuProThrTrpSerThrProValGlnProMet(End)AspProAla -
                                                              STOP


                      P S           P S                            S
           A   HB     ADNNpHPaS     ADNpPaS                        MNDaX
           l   nb     vTlluasui     vTlusui                        bdpuh
           u   lv     aaaaMes9n     aaaMs9n                        oen3o
           1   11     224411161     2241161                        121A3
                      ////  ////    ////                            / //
          GCCCAGCTTGGGGACCCTAGAGGTCCCCTTTTTATTTTGAATTGGGAGATCCCAATTCT
1201      ------------------------------------------------------------ 1260
          CGGGTCGAACCCCTGGGATCTCCAGGGGAAAAAATAAAACTTAACCCTCTAGGGTTAAGA
          AlaGlnLeuGlyAspProArgGlyProLeuPheTyrPheGluLeuGlyAspProAsnSer -
                                        ClaI


                N         A         CT         AHCH A
                l         l         la         laH  l
                a         u         aa         uee  u
                3         1         1          111  1
                                    /               /
          CATGTTTGACAGCTTATCATCGATAAGCTAGCTTTAATGCGGTAGTTTATCACAGTTAAA
1261      ------------------------------------------------------------ 1320
          GTACAAACTGTCGAATAGTAGCTATTCGATCGAAATTACGCCATCAAATAGTGTCAATTT
          HisValEndGlnLeuIleIleAspLysLeuAlaLeuMetArgEndPheIleThrValLys -


                                            H
              B  N                    P     i   H                  HB
              a  l                    o     n   h                  pa
              n  l                    k     p   a                  hn
              1  4                    1     1   1                  11
          TTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCCCTCATCGTCATCCTCGG
1321      ------------------------------------------------------------ 1380
          AACGATTGCGTCAGTCCGTGGCACATACTTTAGATTGTTACGGGAGTAGCAGTAGGAGCC
          LeuLeuThrGlnSerGlyThrValTyrGluIleEndGlnCysAlaHisArgHisProArg -


                  S     BES                                        S S
              N   HHMi  sec                         P       HM  R   HHDNaHc
              l   ana   tor                         o       ps  a   pecuar
              a   elN   NRF                         k       ap  a   api9eF
              4   311   121                         1       21  1   211631
                  //    /                                           ////  /
          CACCGTCACCCTGGATGCTGTAGGCCATAGGCTTGGTTATGCCCGGTACTGCCCGGGCTCTT
1381      -------------------------------------------------------------- 1440
          GTGGCAGTGGGACCTACGACATCCGGTATCCGAACCAATACGGGCCATGACGGGCCCGAGAA
          HisArgHisProGlyCysCysArgHisArgLeuGlyTyrAlaGlyThrAlaGlyProLeu -
```

SUBSTITUTE SHEET

FIG. 11 (cont'd)

```
        E                         S              P    H   S
  M     C                       MB?             a  NM  i HH?
  n     O                       aDa             u  ba  n baa
  i     R                       evN             4  ee  P aeN
  i     V                       3ii             H  ii  i ii:

      GCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATA
1441  ------------------------------------------------------------ 1500
      CGCCCTATAGCAGGTAAGGCTGTCGTAGCGGTCAGTGATACCGCACGACGATCGCGATAT
      AlaGlyTyrArgProPheArgGlnHisArgGlnSerLeuTrpArgAlaAlaSerAlaIle -
```

```
             E                        BH                    P
           1FHM                      sqN                  C Hn
           nsha                      pis                  f au
           Ppat                      1Ap                  r e4
           iiii                      212                  i 3H

      TGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCG
1501  ------------------------------------------------------------ 1560
      ACGCAACTACGTTAAAGATACGCGTGGGCAAGAGCCTCGTGACAGGCTGGCGAAACCGGC
      CysValAspAlaIleSerMetArgThrArgSerArgSerThrValArgProLeuTrpPro -
```

```
    P                           M         T         P    S
    n                           l         a         nMNDaT  N
    u                           a         q         ubdpub  l
    4                           4         i         Doen3a  a
    H                                               2121A1  3

      CCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGAC
1561  ------------------------------------------------------------ 1620
      GGCGGGTCAGGACGAGCGAAGCGATGAACCTCGGTGATAGCTGATGCGCTAGTACCGCTG
      ProProSerProAlaArgPheAlaThrTrpSerHisTyrArgLeuArgAspHisGlyAsp -
```

```
    E                    HM      CXHHHDN?    HMBAGHBHlN
    i                    ps      fgappeaa    psahihbana
    e                    ap      raeahpeN    apnaDaeePr
    i                    21      11321111    2112111211

      CACACCCGTCCTGTGGATTCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCAC
1621  ------------------------------------------------------------ 1680
      GTGTGGGCAGGACACCTAAGAGATGCGGCCTGCGTAGCACCGGCCGTAGTGGCCGCGGTG
      HisThrArgProValAspSerLeuArgArgThrHisArgGlyArgHisHisArgArgHis -
```

```
    S        BAGHBHlNM    H              MNDa      B
    N?       ahihbanal    P              bdpu      BsMON
    la       naDaeePra    b              oen3      apbe
    aN       121112114    l              121A      niop
    41                                             2223

      AGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCA
1681  ------------------------------------------------------------ 1740
      TCCACGCCAACGACCGCGGATATAGCGGCTGTAGTGGCTACCCCTTCTAGCCCGAGCGGT
      ArgCysGlyCysTrpArgLeuTyrArgArgHisHisArgTrpGlyArgSerGlySerPro -
```

```
    B        M   l  HH                    DHNPa     C  HHDNG
    BsM      l   n  ha                    ralsu     f  apscr
    aps      a   P  ae                    aeaa9     r  eapiP
    alp      3   1  12                    23416     l  32111

      CTTCGGGCTCATGAGCGCCTTGTTCGGCGTGGGTATCGTCGCCAGGCCCCGTGGCCGGGG
1741  ------------------------------------------------------------ 1800
      GAAGCCCGAGTACTCGCGGAACAAGCCGCACCCATAGCAGCGGTCCGGGGCACCGGCCCC
      LeuArgAlaHisGluArgLeuPheArgArgGlyTyrGlyGlyArgProArgGlyArgGly -
```

FIG. II(cont'd)

```
                   H        H                                    P        P           BH
                   BAGHBHLNN                                     a        a           aGN        H
                   abihbanal                                     u        u           pls        a
                   naDaaePra                                     4        4           LAp        3
                   121112114                                     H        H           212
                   /   ////                                                           //
    ACTGTTGGGCCGCCATCTCCTTGCACGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCT
1801 ------------------------------------------------------------- 1860
    TGACAACCCGCGGTAGAGGAACGTGCGTGGTAAGGAACGCCGCCGCCACGAGTTGCCGGA
    ThrValGlyArgHisLeuLeuAlaArgThrIleProCysGlyGlyGlyAlaGlnArgPro -
```

```
                                    P                  H        H              S
                                    n                  i        a              a
     B            M                  u                  n        g              a
     b            l                  4                  f        a              H
     v            l                  H                  1        1              l
    CAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGTCCGAT
1861 ------------------------------------------------------------- 1920
    GTTGGATGATGACCCGACGAAGGATTACGTCCTCAGCGTATTCCCTCTCGCAGCAGGCTA
    GlnProThrThrGlyLeuLeuProAsnAlaGlyValAlaEndGlyArgAlaSerSerAsp -
```

```
                                            H  P
       A            HM                       i  nHT              N
       l            ps                       n  uhh              l
       u            ap                       P  Daa              a
       l            21                       1  211              3
                                             /       //
    GCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGT
1921 ------------------------------------------------------------- 1980
    CGGGAACTCTCGGAAGTTGGGTCAGTCGAGGAAGGCCACCCGCGCCCCGTACTGATAGCA
    AlaLeuGluSerLeuGlnProSerGlnLeuLeuProValGlyAlaGlyHisAspTyrArg -
```

```
                                                                P  H
    P                                             N      B  N  HMGN   n  i  HH
    n                      M                       l      a  l  psa    u  n  ha
    u                      b                       a      n  a  apo    4  P  aa
    4                      o                       3      1  4  211    H  1  12
    H                      2                                          //
    CGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCT
1981 ------------------------------------------------------------- 2040
    GCGGCGTGAATACTGACAGAAGAAATAGTACGTTGAGCATCCTGTCCACGGCCGTCGCGA
    ArgArgThrTyrAspCysLeuLeuTyrHisAlaThrArgArgThrGlyAlaGlySerAla -
```

```
                        S                       H  P              S
     BM                 AaS                      i  nHT          MNDa      H
     ba                 vul                      n  uhh          bdpu      a
     vl                 a9n                      P  Daa          oen3      e
     11                 261                      1  211          121A      3
                        //                          //
    CTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCT
2041 ------------------------------------------------------------- 2100
    GACCCAGTAAAAGCCGCTCCTGGCGAAAGCGACCTCGCGCTGCTACTAGCCGGACAGCGA
    LeuGlyHisPheArgArgGlyProLeuSerLeuGluArgAspAspAspArgProValAla -
```

```
                        H                              M        M        AaNS
                        i                              b        a        vuli
                        n                              l        e        a9aa
                        f                              l        3        2641
                        1                                                //
    TGCCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAA
2101 ------------------------------------------------------------- 2160
    ACGGCCATAAGCCTTAGAACGTGCGGGAGCGAGTTCGGAAGCAGTGACCAGGGCGGTGGTT
    CysGlyIleArgAsnLeuAlaArgProArgSerrLeuArgHisTrpSerArgHisGln -
```

32/93

FIG. II(cont'd)

```
          M                      B    N   CE H X    H  T    H    M
          a                      a    i   a au s    u  b     a    a
          p                      p    a   g e a     D  a     a    a
          2                      3    3   H33       2     3

     ACGTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGGCGACGCGCTGGGCTACGT       2220
2161 -------------------------------------------------------------
     TGCAAGCCGCTCTTCGTCCGGTAATAGCGGCCGTACCGCCGCTGCGCGACCCGATGCA

     ThrPheArgArgGluAlaGlyHisTyrArgArgHisGlyGlyArgArgAlaGlyLeuArg -
```

```
          M     N T    a T    H        H        M    H        H P
          b     u b    u b    g        a        o b  a        a
          1     D a    D a    a        e        k o  a        a 2
          1     211    21     1        3        12          2

     CTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTC     2280
2221 -------------------------------------------------------------
     GAACGACCGCAAGCGCTGCGCTCCGACCTACCGGAAGGGGTAATACTAAGAAGAGCGAAG

     LeuAlaGlyValArgAspAlaArgLeuAspGlyLeuProHisTyrAspSerSerArgPhe -
```

```
          S F      S      P       H       B    B E S
          M f n    z      a T     o       a N  a c c
          a a u    a      u b     k       p l  t o r
          p N 4    N      D a     1       H a  N R J
          1 1H     1      21      3       13   121

     CCGCCGGCATCGGGATGCCCGGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCA   2340
2281 --------------------------------------------------------------
     GGCGGCCGTAGCCCTACGGGCCGCAACGTCCGGTACGACAGGTCCGTCCATCTACTGCTGGT

     ArgArgHisArgAspAlaArgValAlaGlyHisAlaValGlnAlaGlyArgEndArgPro -
```

```
          A     HHTDa    aaT          THTDa      AaS
          l     bddpu    uub          abdpu      vui
          u     oeaJ     D4a          qoeaJ      a9a
          1     121A     2H1          1121A      261

     TCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACC    2400
2341 ------------------------------------------------------------
     AGTCCCTGTCGAAGTTCCTAGCGAGCGCCGAGAATGGTCGGATTGAAGCTAGTGACCTGG

     SerGlyThrAlaSerArgIleAlaArgGlySerTyrGlnProAsnPheAspHisTrpThr -
```

```
          N    S          P        B    BN        W        N
          a    HHTDaH     a        g    p i s     l        l
          p    bddpua     u        a    l 1A p    a        a
          B    oeaJe      4        1    1 212     3        3
          2    121A3      H

     GCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGAT    2460
2401 ------------------------------------------------------------
     CGACTAGCAGTGCCGCTAAATACGGCGGAGCCGCTCGTGTACCTTGCCCAACCGTACCTA

     AlaAspArgHisGlyAspLeuCysArgLeuGlyGluHisMetGluArgValGlyMetAsp -
```

```
          H  F H                    P          P        N N HHON
          HAGHBHHHH                 a TH       RT       l 1 pac
          ahihbuaaal                uha        ub       a a apl
          naDae4ePra                Dal        Da       3 4 211
          12111H2114                211        21

     TGTAGGCGGCCGCCCTATACCTTGTCTGCCTCCCCGCCGTTGCGTCGCCGGTGCATGGAGCCG   2520
2461 ----------------------------------------------------------------
     ACATCCGCCGGCGGGGATATGGAACAGACGGAGGGGCGGCAACGCAGCGGCCACGTACCTCGGC

     CysArgArgArgThr IleProCysLeuProProArgValAlaSerArgCysMetGluPro -
```

-

F I G. I I (cont'd)

```
      S S                            P                      H
      a H C          :          M    H X N   n B   N        H    M    l
      u a r          a          B    p s a   u a   l        p    B    B
      9 e P          q          l    a p e   4 B   a        b    l    f
      6 3 1          :          l    2 1 1   H 1   4        l    l    :
      / /                            //
      GGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGA
2521  -----------------------------------------------------------  2580
      CCGGTGGAGCTGGACTTACCTTCGGCCGCCGTGGAGCGATTGCCTAAGTGGTGAGGTTCT
      GlyHisLeuAspLeuAsnGlySerArgArgHisLeuAlaAsnGlyPheThrThrProArg -

      P                                  H          P
      f          M                       l F B H N  s f              H
      l          l                       n s e b s  t l              g
      M          l                       P p m a t  Y M              q
      l          4                       1 1 1 1 1  1 1              l
                                         //
      ATTGGAGCCAATCAATTCTTGCCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAAC
2581  ------------------------------------------------------------  2640
      TAACCTCGGTTAGTTAAGAACGCCTCTTGACACTTACGCGTTTGGTTGGGAACCGTCTTG
      IleGlyAlaAsnGlnPheLeuArgArgThrValAsnAlaGlnThrAsnProTrpGlnAsn -

      P                      P  P  P P  H          S          S
      n T                    a  a  n n S i T H   A B  f     B H D N a  A
      u b                    u  u  u u b a h b   v b  a     c b p d u  l
      D a                    4  4  D 4 v P a a   a v  N     l o n e 3  u
      2 1                    H  H  2 H 1 1 1 1   1 1  1     1 1 1 2 A  1
                                          /                /  ///
      ATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGGGATGATCAG
2641  ------------------------------------------------------------  2700
      TATAGGTAGCGCAGGCGGTAGAGGTCGTCGGCGTGCGCCGCGTAGAGCCCCCTACTAGTC
      IleSerIleAlaSerAlaIleSerSerSerArgThrArgArgIleSerGlyAspAspGln -

      P N P P     P H F P                H          H    M  N B s A   H X N C  S
      n s P P     n i n T H T            n          n    n  a l u p l p a c r  b
      u p v o     u n u h h h            p          p    l  l a 4 H u a p l P  v
      4 B u k     D P D a a l a          b          b    l  1 3 H 1 1 2 1 1 1  1
      //          / / / /                l          l    1              ///
      CTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCCGGAGAC
2701  ------------------------------------------------------------  2760
      GACGGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTG
      LeuProArgAlaPheArgEndEndArgEndLysProLeuThrHisAlaAlaProGlyAsp -

      M          A          S          S          P          H        H P        N
      a          l          f          H X N C     p          g        i n H T    a
      e          u          N          p a c r     o          q        a u h b    p
      3          1          1          a p l P     k          l        P D a a    B
                                       2 1 1 1     1          l        1 2 1 1    2
                                            ///                        //
      GGTCACAGCTTGTCTGTAAGCCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGC
2761  ------------------------------------------------------------  2820
      CCAGTGTCGAACAGACATTCGGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCG
      GlyHisSerLeuSerValSerGlyCysArgGluGlnThrSerProSerGlyArgValSer -

      H P                N    T B M    M                              A
      i H n              l    t B a    a                              c
      n h u              a    b v e    e                              c
      P a 4              3    1 1 3    2                              l
      1 1 H                                                          1
      GGGTGTTGGCCGGGTGTCGGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTA
2821  ------------------------------------------------------------  2880
      CCCACAACCGGCCCACAGCCCCGCGTCGGTACTGGGTCAGTGCATCGCTATCGCCTCACAT
      GlyCysTrpArgValSerGlyArgSerHisAspProValThrEndArgEndArgSerVal -
```

34/93

FIG.11(cont'd)

```
        P              S        R  D      A    BH
        a              a        s  d      p    sçNN
        u              u        a  d      a    p:sd
        4              N        1  1      :    :Ape
        H              1               //      2:21
      TACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGT
2881  ------------------------------------------------------------  2940
      ATGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACA
      TyrTrpLeuAsnTyrAlaAlaSerGluGlnIleValLeuArgValHisHisMetArgCys -

                                    M   H    S              B
                                    b   i    HHf            b
                                    o   n    haa            v
                                    2   P    aeN            1
                                        1    121
      GAAATACCCCACAGATGCGTAAGGAGAAATACCGCATCAGGCGGCTCTTCCGCTTCCTG
2941  ------------------------------------------------------------  3000
      CTTTATGGGGTGTCTACGCATTCCTCTTTATGGCGTAGTCCGCGAGAAGGCGAAGGAGC
      GluIleProHisArgCysValArgArgLysTyrArgIleArgArgSerSerAlaSerSer -

              H    P H              P      P
          M   i    a i    BH        a      a
          b   n    u n    bh        u      u          A
          l   f    u f    va        4      4          l
          1   1    H 1    11        H      H          1
      CTCACTGACTCGCTGCGGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAG
3001  ------------------------------------------------------------  3060
      GAGTGACTGAGCGACGCCGAGCCAGCAAGCCGACGCCGCTCGCCATAGTCGAGTGAGTTTC
      LeuThrAspSerLeuArgSerValValArgLeuArgArgAlaValSerAlaHisSerLys -

                       H                              N
                       i                              Ne
                       n                              1p
                       f                              aH
                       1                              31
      GCGGTAATACGGTTATCCACAGAATCAGCGGGATAACGCAGGAAAGAACATGTGAGCAAAA
3061  ------------------------------------------------------------  3120
      CGCCATTATGCCAATAGGTGTCTTAGTCGCCCTATTGCGTCCTTTCTTGTACACTCGTTTT
      AlaValIleArgLeuSerThrGluSerGlyAspAsnAlaGlyLysAsnMetEndAlaLys -

                  BH S
              H   scHc   M          P  P
              a   toar   1          Ha aT              N
              e   NRaP   a          au uh              i
              3   1234   4          e4 Da              a
                                    3H 21              4
      GGCCAGCAAAGGCCAGGAACCGTAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTC
3121  ------------------------------------------------------------  3180
      CCGGTCGTTTCCGGTCCTTGGCATTTTTCCGGCGCAACGACCGCAAAAGGTATCCGAG
      GlyGlnGlnLysAlaArgAsnArgLysLysAlaAlaLeuLeuAlaPhePheHisArgLeu -

                   S
                   f      T     M          H
                   a      a     b          g
                   N      q     l          a
                   1      1     1          1
      CGCCCCCCTGACGAGCATCACAAAAATCGACGGCTCAAGTCAGAGGTGGCGAAACCCGACA
3181  ------------------------------------------------------------  3240
      GCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCCGAGTTCAGTCTCCACCGCTTTGGGCTGT
      ArgProProAspGluHisHisLysAsnArgArgSerSerGlnArgTrpArgAsnProThr -
```
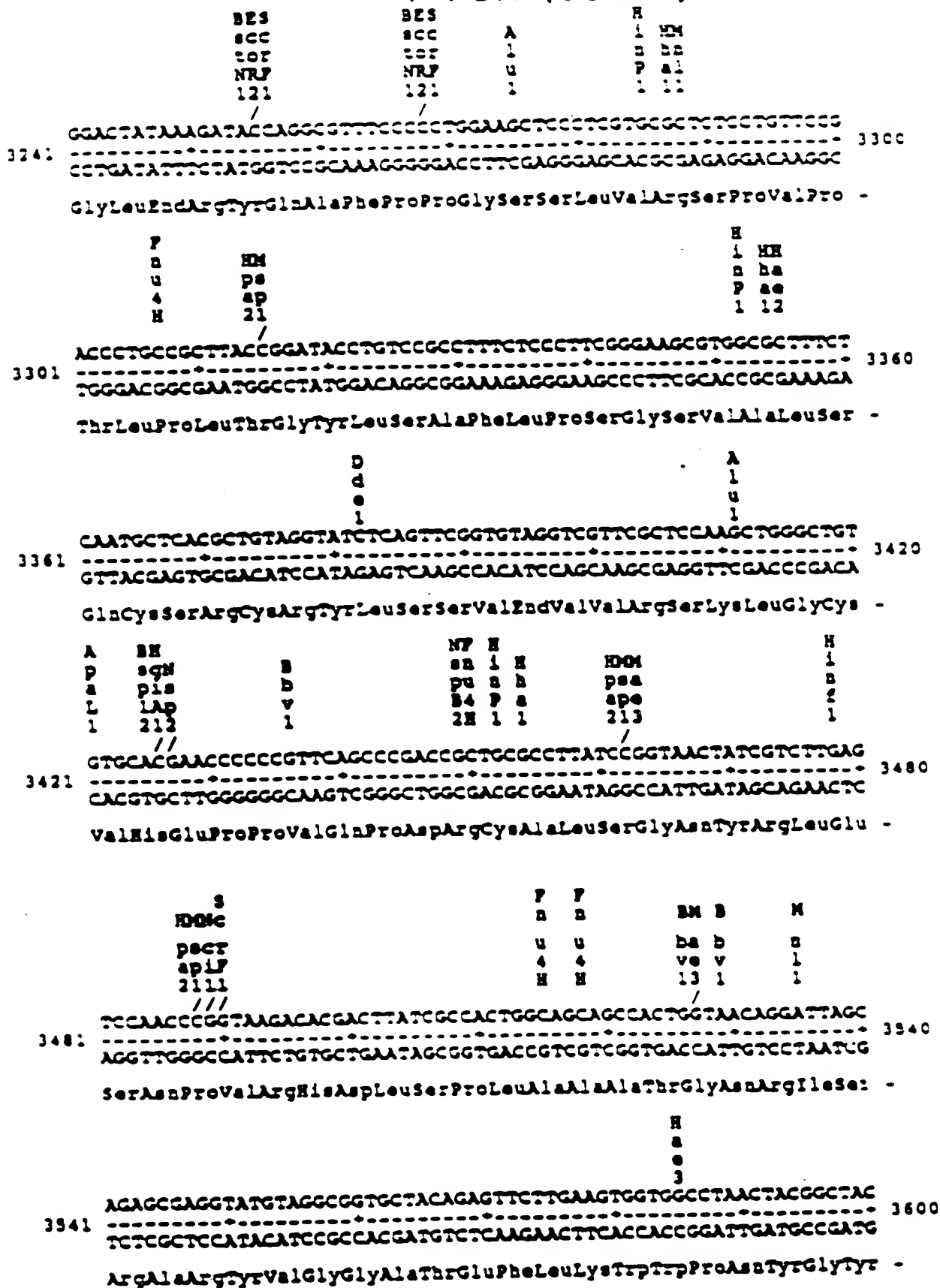
FIG.11(cont'd)

```
         BES        BES           R
         scc        scc        A  i  HM
         tor        tor        l  s  hs
         NRJ        NRJ        u  P  al
         121        121        l  l  ll
           /          /
     GGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGGCTTCCCTGTTCGG
3241 ------------------*-------------------*------------------*     3300
     CCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGC
     GlyLeuLysArgTyrGlnAlaPheProProGlySerSerLeuValArgSerProValPro -

       F
       n          HM                                 H
       u          ps                                 i  HH
       4          sp                                 n  ha
       H          21                                 P  ae
         /                                           l  12
     ACCCTGCCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCCGCTTTCT
3301 ------------------*-------------------*------------------*     3360
     TGGGACGGGCGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGGCGAAAGA
     ThrLeuProLeuThrGlyTyrLeuSerAlaPheLeuProSerGlySerValAlaLeuSer -

                        D                          A
                        d                          l
                        e                          u
                        l                          l
     CAATGCTCACGCTGTAGGTATCTCAGTTCCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGT
3361 ------------------*-------------------*------------------*     3420
     GTTACGAGTGCGACATCCATAGAGTCAAGGCCACATCCAGCAAGCGAGGTTCGACCCGACA
     GlnCysSerArgCysArgTyrLeuSerSerValEndValValArgSerLysLeuGlyCys -

     A     BH                    NF  H              HMM            H
     P     sgN        B          sn  i  H           pss           i
     a     pls        b          pu  s  h           spe           n
     L     LAp        v          s4  P  a           213           f
     l     213        l          2H  l  l             /           l
       //                                            /
     GTGCACGAACCCCCCGTTCAGCCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAG
3421 ------------------*-------------------*------------------*     3480
     CACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGAACTC
     ValHisGluProProValGlnProAspArgCysAlaLeuSerGlyAsnTyrArgLeuGlu -

              S                        F  F
              HMMc                     n  n        HM  B        M
                                       u  u        hs  b        n
              pscr                     4  4        ve  v        l
              aplF                     H  H        13  1        l
              2111                                  /
               ///
     TCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGC
3481 ------------------*-------------------*------------------*     3540
     AGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCG
     SerAsnProValArgHisAspLeuSerProLeuAlaAlaAlaThrGlyAsnArgIleSer -

                                                    H
                                                    a
                                                    e
                                                    3
     AGAGCCGAGGTATGTAGGCCGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTAC
3541 ------------------*-------------------*------------------*     3600
     TCTCGGCTCCATACATCCGGCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATG
     ArgAlaArgTyrValGlyGlyAlaThrGluPheLeuLysTrpTrpProAsnTyrGlyTyr -
```

F I G. II(cont'd)

```
                                    H
                                    i
          M                         n        H                    M
          a                         d        a                    a
          e                         P        e                    e
          9                         l        l                    3

          ACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAGA
3601      ---------------------------------------------------------    3660
          TGATCTTCCTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTCT

          ThrArgArgThrValPheGlyIleCysAlaLeuLeuLysProValThrPheGlyLysArg  -


                         S
               A         MnTDaHN                 N
               l         bdpupa                  s
               u         oen3ap                  P
               1         121A31                  2
                          / / /                   /
          GTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTGC
3661      ---------------------------------------------------------    3720
          CAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGCCACCAAAAAACAACG

          ValGlySerSerEndSerGlyLysGlnThrThrAlaGlySerGlyGlyPhePheValCys  -


               P         P  H       S           S          S
               n         Bn HiT     MnDaX       MnDaX      MnTDMa
               u         bu hah     bdpuh       bdpuh      bdpbu
               4         vDaPa      oen3o       oen3o      oen3
               H         12111      12 1A2      121A2      1212A
                          / //      / //        / //        / //
          AAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACG
3721      ---------------------------------------------------------    3780
          TTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGATGC

          LysGlnGlnIleThrArgArgLysLysGlySerGlnGluAspProLeuIlePheSerThr  -


                    D     H            M              N        M
                    d     g            a              l        b
                    e     l            e              a        o
                    1     1            2              3        2
          GGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCA
3781      ---------------------------------------------------------    3840
          CCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGT

          GlySerAspAlaGlnTrpAsnGluAsnSerArgEndGlyIleLeuValMetArgLeuSer  -


               S    MnDaX      M   MnDaX       BMDMaX    MnTDa      A
               H    bdpuh      a   bdpuh       qbpduh    bdpu       l
               P    oen3o      e   oen3o       1one3o    oen3       u
               l    121A2      l   121A2       2112A2    121A       1
                    / //       /   / //        / ///     / //
          AAAAGGATCTTCACCTAGATCCTTTTCAGATCTCCCGATCTTAGCTGTCTTGGTTTGCC
3841      ---------------------------------------------------------    3900
          TTTTCCTAGAAGTGGATCTAGGAAAAGTCTAGAGGGCTAGAAATCGACAGAACCAAACGG

          LysArgIlePheThrEndIleLeuPheArgSerProAspLeuEndLeuSerTrpPheAla  -


                    H
                    i     H                                    D    MR
                    n     h                                    d    ns
                    P     a                                    e    la
                    l     1                                    1    11
                                                                     /
          CAAAGCCGCATTGCATAATCTTTCAGGCGTTATGCGTTGTTCCATACAACCTCCTTAGTACA
3901      ---------------------------------------------------------    3960
          GTTTCGGCGTAACGTATTAGAAAGTCCGCAATACGCAACAAGGTATGTTGGAGGAATCATGT

          GlnSerAlaLeuHisAsnLeuSerGlyLeuCysValValProTyrAsnLeuLeuSerThr  -
```

FIG.11(cont'd)

```
        N                                          H
       N s H              M                        i
       1 P P              s                        n
       4 H b              p                        c
       3 1 1              1                        2
        /
       TGCAACCATTATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGATATTTATC
3961   ----------------*----------------*----------------*---------*   4020
       ACGTTGGTAATAGTGGCGGTCTCCATTTTATCAGTTGTGCGTGCCACAATCTATAAATAG
       CysAsnHisTyrHisArgGlnArgEndAsnSerGlnHisAlaArgCysEndIlePheIle -


                              B H                              F
                HM            s G N                     M       n   A
                p a           p i s                     n       u   1
                b o           1 A p                     1       4   u
                1 3           2 1 3                     1       H   1
                                      //
       CCTTGCGGTGATAGATTTAACGTATGAGCACAAAAAGAAACCATTAACACAAGAGCAGC
4021   ----------------*----------------*----------------*---------*   4080
       GGAACGCCACTATCTAAATTGCATACTCGTGTTTTTTCTTTGGTAATTGTGTTCTCGTCG
       ProCysGlyAspArgPheAsnValEndAlaGlnLysArgAsnHisEndHisLysSerSer -


            B    M    H
            b    a    g
            v    e    a
            1    2    1
       TTGAGGACGCACGTCGCCCTTAAAGCAATTTATGAAAAAAGAAAATGAACTTGGCTTAT
4081   ----------------*----------------*----------------*---------*   4140
       AACTCCTGCGTGCAGCGGGAATTTCGTTAAATACTTTTTTCTTTTTACTTGAACCGAATA
       LeuArgThrHisValAlaLeuLysGlnPheMetLysLysArgLysMetAsnLeuAlaTyr -


       B Z S   H                                       F
       s C C   i                                       o
       t o r   n                                       k
       N R P   c                                       1
       1 2 1   1
        /
       CCCAGGAATCTGTCGCAGACAAGATGGGGATCGGGCAGTCAGGCGTTGGTGCTTTATTA
4141   ----------------*----------------*----------------*---------*   4200
       GGGTCCTTAGACAGCGTCTGTTCTACCCCTACCCCGTCAGTCCGCAACCACGAAATAAAT
       ProArgAsnLeuSerGlnThrArgTrpGlyTrpGlySerGlnAlaLeuValLeuTyrLeu -


       S             F
       2 N           n
       a a           u
       N 1           4
       1 1           H
       ATGGCATCAATGCATTAAATGCTTATAACGCCGCATTGCTTACAAAAATTCTCAAAGTTA
4201   ----------------*----------------*----------------*---------*   4260
       TACCGTAGTTACGTAATTTACGAATATTGCGGCGTAACGAATGTTTTTAAGAGTTTCAAT
       MetAlaSerMetHisEndMetLeuIleThrProHisCysLeuGlnLysPheSerLysLeu -


            M
            b
            2
       GCGTTGAAGAATTTAGCCCCTTCAATCGCCAGAGAAATCTACGAGATGTATGAAGCCGTTA
4261   ----------------*----------------*----------------*---------*   4320
       CGCAACTTCTTAAATCGGGAAGTTAGCGGTCTCTTTAGATGCTCTACATACTTCGGCAAT
       AlaLeuLysAsnLeuAlaLeuGlnSerProGluLysSerThrArgCysMetLysArgLeu -
```

SUBSTITUTE SHEET

FIG. II(cont'd)

```
           P                M        D  B                        R                        N
           t                M        d  b                        e                        l
           u                e        e  v                        a                        a
           H                3        1  1                        1                        3
     GTATGCAGCCGTCACTTAGAAGTGAGTATGAGTACCCTGTTTTTCTCATGTTCAGGCAG   4380
4321 ----------------------------------------------------------
     CATACGTCGGCAGTGAATCTTCACTCATACTCATGGGACAAAAAGAGTACAAGTCCGTC
     ValCysSerArgHisLeuGluValSerMetSerThrLeuPhePheLeuMetPheArgGln -
```

```
     H                E        S                        H
     P          DFi  AD       f                        P
     h          don  ld       a                        h
     1          ekd  ue       N                        1
                113  11
     GGATGTTCTCACCTAAGCTTAGAACCTTTACCAAAGGTCATGCGGAGAGATCGGTAAGCA   4440
4381 ----------------------------------------------------------
     CCTACAAGAGTGGATTCGAATCTTGGAAATGGTTTCCACTACGCCTCTCTAGCCATTCGT
     GlyCysSerHisLeuSerLeuGluProLeuProLysValMetArgArgAspGlyEndAla -
```

```
                    H              BM                           N
                    i              sn                           l
                    n              al                           a
                    f              11                           3
                    1
     CAACCAAAAAGCCAGTGATTCTGCATTCTGGCTTGAGGTTGAAGGTAATTCCATGACCG    4500
4441 ----------------------------------------------------------
     GTTGGTTTTTCGGTCACTAAGACGTAAGACCGAACTCCAACTTCCATTAAGGTACTGGC
     GlnProLysLysProValIleLeuHisSerGlyLeuArgLeuLysValIleProEndPro -
```

```
               N        H    A                        H        D
               l        i    l                        i        d
               a        n    u                        n        e
               4        d3   1                        c2       1
     CACCAACAGGCTCCAAGCCAGCTTTCCTGACGGAATGTTAATTCTCGTTGACCCTGAGC    4560
4501 ----------------------------------------------------------
     GTGGTTGTCCGAGGTTCGGTTCGAAAGGACTGCCTTACAATTAAGAGCAACTGGGACTCG
     HisGlnGlnAlaProSerGlnAlaPheLeuThrGluCysEndPheSerLeuThrLeuSer -
```

```
             A
             Tcc                  H                        H
             tcr                  p                        h
             NR7                  h                        1
             121                  1
     AGGCTGTTGAGCCAGGTGATTTCTGCATAGCCAGACTTGGGGGTGATGAGTTTACCTTCA    4620
4561 ----------------------------------------------------------
     TCCGACAACTCGGTCCACTAAAGACGTATCGGTCTGAACCCCCACTACTCAAATGGAAGT
     ArgLeuLeuSerGlnValIleSerAlaEndProAspLeuGlyValMetSerLeuProSer -
```

```
                                                             R
                                                             s
                                                             a
                                                             1
     AGAAACTAATTAGGGATAGCGGTCAGGTGTTTTTACAACCACTAAACCCACAGTACCCAA    4680
4621 ----------------------------------------------------------
     TCTTTGATTAATCCCTATCGCCAGTCCACAAAAATGTTGGTGATTTGGGTGTCATGGGTT
     ArgAsnEndLeuGlyIleAlaValArgCysPheTyrAsnHisEndThrHisSerThrGln -
```

FIG. II (cont'd)

```
         S
       KNDa              N                                        M              H
       bdpu              i                                        a              a
       oen3              a                                        e              e
       12LA              3                                        :              3
       / /               /                                        :              :

       TCATCCATGCAATGAGAGTTGTTCCGTTGTGGGAAAGTTATCGCTAGTCAGTGGCCTG
4681   --------------------------------------------*-------------------  4740
       ACTAGGGTACGTTACTCTCAACAAGGCAACACCCCTTTCAATAGCGATCAGTCACCGGAC

       EndSerHisAlaMetArgValValProLeuTrpGlyLysLeuSerLeuValSerGlyLeu  -
```

```
              M         M    KNDa                       H
              a         b    bdpu               i        H          A              M
              e         o    oen3               a        h          l              b
              3         2    12LA               P        a          u              o
                             / /                1        1          1              2

       AAGAGACGTTTGGCTGATCGGCAAGGTGTTCTCGGTCGGCGGCATAGCTGATAACAATTGAC
4741   ---------------------------------------------*-------------------  4800
       TTCTCTGCAAACCGACTAGCCGTTCCACAAGAGCCAGCCGCCGTATCGACTATTGTTAACTC

       LysArgArgLeuAlaAspArgGlnGlyValLeuValGlyAlaEndLeuIleThrIleGlu  -
```

```
         H                  PS  P              BM     HM          KNDaX
         iS                 nf  aKP            ba     pe          pbdpuh
         nb                 ua  uga            vl     ap          hoen3o
         fv                 4M  4at            ll     21          112LA2
         ll                 Hl  Hll                               / /  //

       CAAGAATCTTCATCGGGGCTGCAGCCCACGATGCGTCCGGCGTAGAGGATCTCTCACCTA
4801   --------------------------------------------*-------------------  4860
       GTTCTTAGAAGTAGCCCCGACGTCGGGTGCTACGCAGGCCGCATCTCCTAGAGAGTGGAT

       GlnGluSerSerSerGlyLeuGlnProThrMetArgProAlaEndArgIleSerHisLeu  -

       CCAAACAATGCCCCCCTGCAAAAAATAAATTCATATAAAAACATACAGATAACCATCTG
4861   --------------------------------------------*-------------------  4920
       GGTTTGTTACGGGGGGACGTTTTTTATTTAAGTATATTTTTGTATGTCTATTGGTAGAC

       ProAsnAsnAlaProLeuGlnLysIleAsnSerTyrLysLysHisThrAspAsnHisLeu  -
```

```
                           H                                    D      BH
              E            i                                    d      Hsg
              P            n                                    e      PP1
              h            c                                           h1A
                                                                       / 
              1            2                                    1      121

       CGGTGATAAATTATCTCTGGCGGTCTTGACATAAATACCACTGGCGGTGATACTGAGCAC
4921   --------------------------------------------*-------------------  4980
       GCCACTATTTAATAGAGACCGCCACAACTGTATTTATGGTGACCGCCACTATGACTCGTG

       ArgEndEndIleIleSerGlyGlyValAspIleAsnThrThrGlyGlyAspThrGluHis  -
```

```
              M               E         M    M          HH
              a               g         l    a          TP
              P               a         a    e          ah
              1               1         3    3          11
              /                                         /

       ATCAGCAGGACGCCACTGACCACCATGAAGGTCACGCTCTTAAAATTAAGCCCTGAAGAAG
4981   --------------------------------------------*-------------------  5040
       TAGTCGTCCTGCGGTGACTGGTGGTACTTCCACTGCGAGAATTTTAATTCGGGACTTCTTC
```

FIG. II(cont'd)

```
F
a  BM              b
u  sb              b
4  no              v
H  13              1
GGCAGCATTCAAAGCAGAAGGCCTTTGGGGTGTGTGATACGAAACGAAGCATT
5041 --------+---------+---------+---------+---------+---------  5092
CCGTCGTAAGTTTCGTCTTCCGAAACCCCACACACTATGCTTTGCTTCGTAA

GlySerIleGlnSerArgArgLeuTrpGlyValEndTyrGluThrLysHis??? -
```

Enzymes that do cut:

Acc1	Aha2	Afl2	Alu1	ApaL1	Ava1	Ava2	Ban1	Ban2	Bbe1	Bbv1
Bcl1	Bgl1	Bgl2	Bcm1	Bsp12	BspM1	BstE2	BstN1	BstX1	Cfr1	Cla1
Dde1	Dpa1	Dra2	Eag1	Eco8	EcoK	EcoR1	EcoR2	EcoRV	FnuD2	Fnu4H
Fok1	Fsp1	Hae2	Hae3	Hga1	HgiA1	HgiD1	Hha1	Hinc2	Hind3	Hinf1
HinP1	Hpa2	Hph1	Mae1	Mae2	Mae3	Mbo1	Mbo2	Mnl1	Msp1	Mst1
Mst2	Nae1	Nar1	Nci1	Nde1	Nde2	Nhe1	Nla3	Nla4	Nru1	Nsi1
Nsp2	NspB2	NspH1	PflM1	PpuM1	Pss1	Pst1	Pvu2	Rsa1	Sac1	Sal1
Sau3A	Sau96	ScrF1	SfaN1	Sin1	Snt1	Stu1	Sty1	Taq1	Tha1	Tth1
Xho2	Xma3									

Enzymes that do not cut:

Aat2	Aha3	Apa1	Asp70	Asp71	Asu2	Avr2	Bal1	BamH1	BspM2	BssH2
Dra1	Dra3	Esp1	Hpa1	Kpn1	Mlu1	Nco1	Not1	PaeR7	Pvu1	Rsr2
Sac2	Sca1	Sfi1	Sma1	SnaB1	Spe1	Sph1	Ssp1	Sst2	Xba1	Xho1
Xma1	Xmn1	Xor2								

F I G. 12
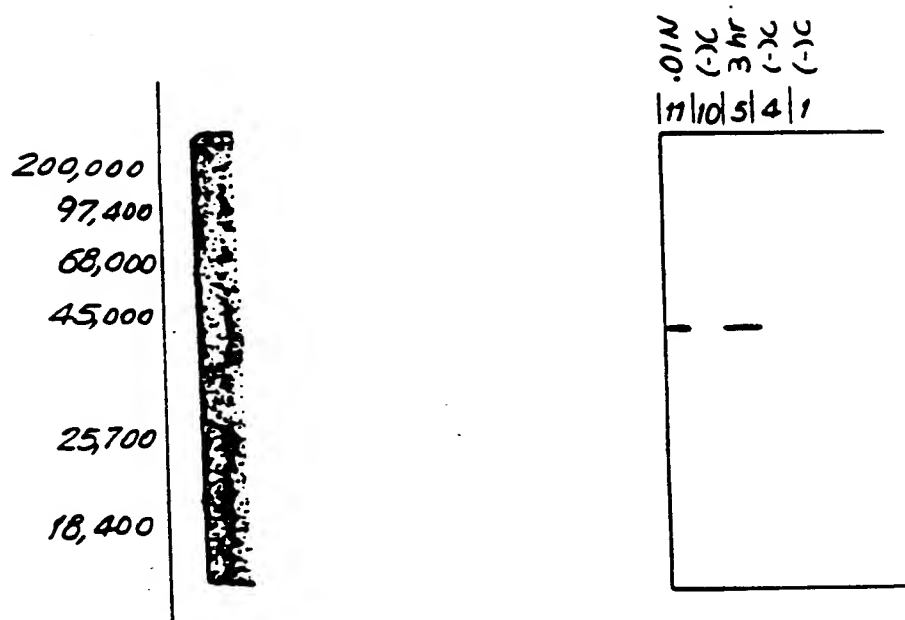
lane 1 = Pre-induced
lane 4 = Uninduced
lane 5 = 3 hr. post-induction
lane 11 = overnight post-induction (~16 hr)
MW = Molecular wt. markers

42/93

FIG. 13

BG 368
(4.75 kb)

SV40 (StuI/SstII)
AD
MLP
XhoI
HindIII
EcoRI
BgII
.34
.04
BamHI
SphI
SalI

(1) EcoRI
(2) BgII
(3) Klenow
(4) Nucleotides
(5) Religate

EcoRI SV40 (StuI/SstII)
AD
MLP
XhoI
BgII
HindIII BamHI
XhoI
EcoRI
SmaI
ScaI
BgII
.34
.66
.04
.85
BamHI
SphI
SalI

BG 312
(4.75 kb)

Amp^R

ori

FIG.14A

pBG 377

EcoRI/MstII, Isolate 1.3 kbp
fragment

pBG 368·EcoRI/BglII

p170-2·MstII/XhoII
Isolate 0.32 kbp
fragment

p170-2·MstII/BglII
Isolate 1.2 Kbp
fragment

Ligase

Ligase

pBG 378
7.25 Kbp
rf1 T4

Amp^r

HindIII

T4

SalI

ori

BglII

PA

p185-23·MstII/BglII, Isolate
0.1 Kbp fragment

Ligase

pBG 379
6.05 Kbp
Δ·3' UTS
rf1 T4

Amp^r

HindIII

T4

SalI

ori

(BglII/XhoII)

PA

pBG 380
6.15 Kbp
rsT4.2

Amp^r

HindIII

rsT4

SalI

ori

BglII

PA

SUBSTITUTE SHEET

FIG. 14B

pBG 377

      EcoRI/MstII, Isolate
      1.3 Kbp fragment

      pBG 368 · EcoRI/Bgl II

      p186·6 · Mst II/Bgl II, Isolate
      0.11 kbp fragment

Amp^r

pBG 381
rST4.3
6.2 Kbp

rST4

Hind III

Sal I

Bgl II

ori

PA

45/93

FIG.14C

pBG391
:BG368 backbone
:soluble T4#3
:AA #3 = LYS

FIG.15

bg381.seq   Length: 6151

```
   1   GAATTAATTC CAGCTTGCTG TGGAATGTGT GTCAGTTAGG GTGTGGAAAG

  51   TCCCCAGGCT CCCCAGCAGG CAGAAGTATG CAAAGCATGC ATCTCAATTA

 101   GTCAGCAACC AGGTGTGGAA AGTCCCCAGG CTCCCCAGCA GGCAGAAGTA

 151   TGCAAAGCAT GCATCTCAAT TAGTCAGCAA CCATAGTCCC GCCCCTAACT

 201   CCGCCCATCC CGCCCCTAAC TCCGCCCAGT TCCGCCCATT CTCCGCCCCA

 251   TGGCTGACTA ATTTTTTTTA TTTATGCAGA GGCCGAGGCC GCCTCGGCCT

 301   CTGAGCTATT CCAGAAGTAG TGAGGAGGCT TTTTTGGAGG GGTCCTCCTC

 351   GTATAGAAAC TCGGACCACT CTGAGACGAA GGCTCGCGTC CAGGCCAGCA

 401   CGAAGGAGGC TAAGTGGGAG GGGTAGCGGT CGTTGTCCAC TAGGGGGTCC

 451   ACTCGCTCCA GGGTGTGAAG ACACATGTCG CCCTCTTCGG CATCAAGGAA

 501   GGTGATTGGT TTATAGGTGT AGGCCACGTG ACCGGGTGTT CCTGAAGGGG

 551   GGCTATAAAA GGGGGTGGGG GCGCGTTCGT CCTCACTCTC TTCCGCATCG

 601   CTGTCTGCGA GGGCCAGCTG TTGGGCTCGC GGTTGAGGAC AAACTCTTCG

 651   CGGTCTTTCC AGTACTCTTG GATCGGAAAC CCGTCGGCCT CCGAACGGTA

 701   CTCCGCCACC GAGGGACCTG AGCGAGTCCG CATCGACCGG ATCGGAAAAC

 751   CTCTCGAGAA AGGCGTCTAA CCAGTCACAG TCGCAAGGTA GGCTGAGCAC

 801   CGTGGCGGGC GGCAGCGGGT GGCGGTCGGG GTTGTTTCTG GCGGAGGTGC

 851   TGCTGATGAT GTAATTAAAG TAGGCGGTCT TGAGACGGCG GATGGTCGAG

 901   GTGAGGTGTG GCAGGCTTGA GATCGATCTG GCCATACACT TGAGTGACAA

 951   TGACATCCAC TTTGCCTTTC TCTCCACAGG TGTCCACTCC CAGGTCCAAC

1001   TGGATCCAAG CTTCGACTCG AGGAATTCCC CGAAGGAACA AAGCACCCTC

1051   CCCACTGGGC TCCTGGTTGC AGAGCTCCAA GTCCTCACAC AGATACGCCT

1101   GTTTGAGAAG CAGCGGGCAA GAAAGACGCA AGCCCAGAGG CCCTGCCATT

1151   TCTGTGGGCT CAGGTCCCTA CTGGCTCAGG CCCCTGCCTC CCTCGGCAAG

1201   GCCACAATGA ACCGGGGAGT CCCTTTTAGG CACTTGCTTC TGGTGCTGCA

1251   ACTGGCGCTC CTCCCAGCAG CCACTCAGGG AAAGAAAGTG GTGCTGGGCA

1301   AAAAAGGGGA TACAGTGGAA CTGACCTGTA CAGCTTCCCA GAAGAAGAGC

1351   ATACAATTCC ACTGGAAAAA CTCCAACCAG ATAAAGATTC TGGGAAATCA

1401   GGGCTCCTTC TTAACTAAAG GTCCATCCAA GCTGAATGAT CGCGCTGACT
```

FIG. 15(cont'd)

```
1451  CAAGAAGAAG CTTGTGGGAC CAAGGAAACT TTCCCCTGAT CATCAAGAAT
1501  CTTAAGATAG AAGACTCAGA TACTTACATC TGTGAAGTGG AGGACCAGAA
1551  GGAGGAGGTG CAATTGCTAG TGTTCGGATT GACTGCCAAC TCTGACACCC
1601  ACCTGCTTCA GGGGCAGAGC CTGACCCTGA CCTTGGAGAG CCCCCCTGGT
1651  AGTAGCCCCT CAGTGCAATG TAGGAGTCCA AGGGGTAAAA ACATACAGGG
1701  GGGGAAGACC CTCTCCGTGT CTCAGCTGGA GCTCCAGGAT AGTGGCACCT
1751  GGACATGCAC TGTCTTGCAG AACCAGAAGA AGGTGGAGTT CAAAATAGAC
1801  ATCGTGGTGC TAGCTTTCCA GAAGGCCTCC AGCATAGTCT ATAAGAAAGA
1851  GGGGGAACAG GTGGAGTTCT CCTTCCCACT CGCCTTTACA GTTGAAAAGC
1901  TGACGGGCAG TGGCGAGCTG TGGTGGCAGG CGGAGAGGGC TTCCTCCTCC
1951  AAGTCTTGGA TCACCTTTGA CCTGAAGAAC AAGGAAGTGT CTGTAAAACG
2001  GGTTACCCAG GACCCTAAGC TCCAGATGGG CAAGAAGCTC CCGCTCCACC
2051  TCACCCTGCC CCAGGCCTTG CCTCAGTATG CTGGCTCTGG AAACCTCACC
2101  CTGGCCCTTG AAGCGAAAAC AGGAAAGTTG CATCAGGAAG TGAACCTGGT
2151  GGTGATGAGA GCCACTCAGC TCCAGAAAAA TTTGACCTGT GAGGTGTGGG
2201  GACCCACCTC CCCTAAGCTG ATGCTGAGTT TGAAACTGGA GAACAAGGAG
2251  GCAAAGGTCT CGAAGCGGGA GAAGGCGGTG TGGGTGCTGA ACCCTGAGGC
2301  GGGGATGTGG CAGTGTCTGC TGAGTGACTC GGGACAGGTC CTGCTGGAAT
2351  CCAACATCAA GGTTCTGCCC ACATGGTCGA CCCCGGTGCA GCCAATGGCC
2401  CTGATTTGAG ATCTTTGTGA AGGAACCTTA CTTCTGTGGT GTGACATAAT
2451  TGGACAAACT ACCTACAGAG ATTTAAAGCT CTAAGGTAAA TATAAAATTT
2501  TTAAGTGTAT AATGTGTTAA ACTACTGATT CTAATTGTTT GTGTATTTTA
2551  GATTCCAACC TATGGAACTG ATGAATGGGA GCAGTGGTGG AATGCCTTTA
2601  ATGAGGAAAA CCTGTTTTGC TCAGAAGAAA TGCCATCTAG TGATGATGAG
2651  GCTACTGCTG ACTCTCAACA TTCTACTCCT CCAAAAAAGA AGAGAAAGGT
2701  AGAAGACCCC AAGGACTTTC CTTCAGAATT GCTAAGTTTT TTGAGTCATG
2751  CTGTGTTTAG TAATAGAACT CTTGCTTGCT TTGCTATTTA CACCACAAAG
2801  GAAAAAGCTG CACTGCTATA CAAGAAAATT ATGGAAAAAT ATTCTGTAAC
2851  CTTTATAAGT AGGCATAACA GTTATAATCA TAACATACTG TTTTTTCTTA
2901  CTCCACACAG GCATAGAGTG TCTGCTATTA ATAACTATGC TCAAAAATTG
2951  TGTACCTTTA GCTTTTTAAT TTGTAAAGGG GTTAATAAGG AATATTTGAT
3001  GTATAGTGCC TTGACTAGAG ATCATAATCA GCCATACCAC ATTTGTAGAG
3051  GTTTTACTTG CTTTAAAAAA CCTCCCACAC CTCCCCCTGA ACCTGAAACA
```

48/93

FIG. 15(cont'd)

```
3101  TAAAATGAAT  GCAATTGTTG  TTGTTAACTT  GTTTATTGCA  GCTTATAATG
3151  GTTACAAATA  AAGCAATAGC  ATCACAAATT  TCACAAATAA  AGCATTTTTT
3201  TCACTGCATT  CTAGTTGTGG  TTTGTCCAAA  CTCATCAATG  TATCTTATCA
3251  TGTCTGGATC  CTCTACGCCG  GACGCATCGT  GGCCGGCATC  ACCGGCGCCA
3301  CAGGTGCGGT  TGCTGGCGCC  TATATCGCCG  ACATCACCGA  TGGGGAAGAT
3351  CGGGCTCGCC  ACTTCGGGCT  CATGAGCGCT  TGTTTCGGCG  TGGGTATGGT
3401  GGCAGGCCCG  TGGCCGGGGG  ACTGTTGGGC  GCCATCTCCT  TGCATGCACC
3451  ATTCCTTGCG  GCGGCGGTGC  TCAACGGCCT  CAACCTACTA  CTGGGCTGCT
3501  TCCTAATGCA  GGAGTCGCAT  AAGGGAGAGC  GTCGACCGAT  GCCCTTGAGA
3551  GCCTTCAACC  CAGTCAGCTC  CTTCCGGTGG  GCGCGGGGCA  TGACTATCGT
3601  CGCCGCACTT  ATGACTGTCT  TCTTTATCAT  GCAACTCGTA  GGACAGGTGC
3651  CGGCAGCGCT  CTGGGTCATT  TTCGGCGAGG  ACCGCTTTCG  CTGGAGCGCG
3701  ACGATGATCG  GCCTGTCGCT  TGCGGTATTC  GGAATCTTGC  ACGCCCTCGC
3751  TCAAGCCTTC  GTCACTGGTC  CCGCCACCAA  ACGTTTCGGC  GAGAAGCAGG
3801  CCATTATCGC  CGGCATGGCG  GCCGACGCGC  TGGGCTACGT  CTTGCTGGCG
3851  TTCGCGACGC  GAGGCTGGAT  GGCCTTCCCC  ATTATGATTC  TTCTCGCTTC
3901  CGGCGGCATC  GGGATGCCCG  CGTTGCAGGC  CATGCTGTCC  AGGCAGGTAG
3951  ATGACGACCA  TCAGGGACAG  CTTCAAGGAT  CGCTCGCGGC  TCTTACCAGC
4001  CTAACTTCGA  TCACTGGACC  GCTGATCGTC  ACGGCGATTT  ATGCCGCCTC
4051  GGCGAGCACA  TGGAACGGGT  TGGCATGGAT  TGTAGGCGCC  GCCCTATACC
4101  TTGTCTGCCT  CCCCGCGTTG  CGTCGCGGTG  CATGGAGCCG  GGCCACCTCG
4151  ACCTGAATGG  AAGCCGGCGG  CACCTCGCTA  ACGGATTCAC  CACTCCAAGA
4201  ATTGGAGCCA  ATCAATTCTT  GCGGAGAACT  GTGAATGCGC  AAACCAACCC
4251  TTGGCAGAAC  ATATCCATCG  CGTCCGCCAT  CTCCAGCAGC  CGCACGCGGC
4301  GCATCTCGGG  CCGCGTTGCT  GGCGTTTTTC  CATAGGCTCC  GCCCCCCTGA
4351  CGAGCATCAC  AAAAATCGAC  GCTCAAGTCA  GAGGTGGCGA  AACCCGACAG
4401  GACTATAAAG  ATACCAGGCG  TTTCCCCCTG  GAAGCTCCCT  CGTGCGCTCT
4451  CCTGTTCCGA  CCCTGCCGCT  TACCGGATAC  CTGTCCGCCT  TTCTCCCTTC
4501  GGGAAGCGTG  GCGCTTTCTC  AATGCTCACG  CTGTAGGTAT  CTCAGTTCGG
4551  TGTAGGTCGT  TCGCTCCAAG  CTGGGCTGTG  TGCACGAACC  CCCCGTTCAG
4601  CCCGACCGCT  GCGCCTTATC  CGGTAACTAT  CGTCTTGAGT  CCAACCCGGT
4651  AAGACACGAC  TTATCGCCAC  TGGCAGCAGC  CACTGGTAAC  AGGATTAGCA
4701  GAGCGAGGTA  TGTAGGCGGT  GCTACAGAGT  TCTTGAAGTG  GTGGCCTAAC
```

49/93

FIG. 15(cont'd)

```
475:   TACGGCTACA CTAGAAGGAC AGTATTTGGT ATCTGCGCTC TGCTGAAGCC

4801   AGTTACCTTC GGAAAAAGAG TTGGTAGCTC TTGATCCGGC AAACAAACCA

4851   CCGCTGGTAG CGGTGGTTTT TTTGTTTGCA AGCAGCAGAT TACGCGCAGA

4901   AAAAAAGGAT CTCAAGAAGA TCCTTTGATC TTTTCTACGG GGTCTGACGC

4951   TCAGTGGAAC GAAAACTCAC GTTAAGGGAT TTTGGTCATG AGATTATCAA

5001   AAAGGATCTT CACCTAGATC CTTTTAAATT AAAAATGAAG TTTTAAATCA

5051   ATCTAAAGTA TATATGAGTA AACTTGGTCT GACAGTTACC AATGCTTAAT

5101   CAGTGAGGCA CCTATCTCAG CGATCTGTCT ATTTCGTTCA TCCATAGTTG

5151   CCTGACTCCC CGTCGTGTAG ATAACTACGA TACGGGAGGG CTTACCATCT

5201   GGCCCCAGTG CTGCAATGAT ACCGCGAGAC CCACGCTCAC CGGCTCCAGA

5251   TTTATCAGCA ATAAACCAGC CAGCCGGAAG GGCCGAGCGC AGAAGTGGTC

5301   CTGCAACTTT ATCCGCCTCC ATCCAGTCTA TTAATTGTTG CCGGGAAGCT

5351   AGAGTAAGTA GTTCGCCAGT TAATAGTTTG CGCAACGTTG TTGCCATTGC

5401   TGCAGGCATC GTGGTGTCAC GCTCGTCGTT TGGTATGGCT TCATTCAGCT

5451   CCGGTTCCCA ACGATCAAGG CGAGTTACAT GATCCCCCAT GTTGTGCAAA

5501   AAAGCGGTTA GCTCCTTCGG TCCTCCGATC GTTGTCAGAA GTAAGTTGGC

5551   CGCAGTGTTA TCACTCATGG TTATGGCAGC ACTGCATAAT TCTCTTACTG

5601   TCATGCCATC CGTAAGATGC TTTTCTGTGA CTGGTGAGTA CTCAACCAAG

5651   TCATTCTGAG AATAGTGTAT GCGGCGACCG AGTTGCTCTT GCCCGGCGTC

5701   AACACGGGAT AATACCGCGC CACATAGCAG AACTTTAAAA GTGCTCATCA

5751   TTGGAAAACG TTCTTCGGGG CGAAAACTCT CAAGGATCTT ACCGCTGTTG

5801   AGATCCAGTT CGATGTAACC CACTCGTGCA CCCAACTGAT CTTCAGCATC

5851   TTTTACTTTC ACCAGCGTTT CTGGGTGAGC AAAAACAGGA AGGCAAAATG

5901   CCGCAAAAAA GGGAATAAGG GCGACACGGA AATGTTGAAT ACTCATACTC

5951   TTCCTTTTTC AATATTATTG AAGCATTTAT CAGGGTTATT GTCTCATGAG

6001   CGGATACATA TTTGAATGTA TTTAGAAAAA TAAACAAATA GGGGTTCCGC

6051   GCACATTTCC CCGAAAAGTG CCACCTGACG TCTAAGAAAC CATTATTATC

6101   ATGACATTAA CCTATAAAAA TAGGCGTATC ACGAGGCCCT TTCGTCTTCA

6151   A
```

pBG392
:BG368 backbon
:soluble T4#7   *FIG.16*
:AA #3 = LYS
:182AA+6AA
:from 203-5

        bg392.sec Length: 6149

     1 GAATTAATTC CAGCTTGCTG TGGAATGTGT GTCAGTTAGG GTGTGGAAAG

    51 TCCCCAGGCT CCCCAGCAGG CAGAAGTATG CAAAGCATGC ATCTCAATTA

   101 GTCAGCAACC AGGTGTGGAA AGTCCCCAGG CTCCCCAGCA GGCAGAAGTA

   151 TGCAAAGCAT GCATCTCAAT TAGTCAGCAA CCATAGTCCC GCCCCTAACT

   201 CCGCCCATCC CGCCCCTAAC TCCGCCCAGT TCCGCCCATT CTCCGCCCCA

   251 TGGCTGACTA ATTTTTTTTA TTTATGCAGA GGCCGAGGCC GCCTCGGCCT

   301 CTGAGCTATT CCAGAAGTAG TGAGGAGGCT TTTTTGGAGG GGTCCTCCTC

   351 GTATAGAAAC TCGGACCACT CTGAGACGAA GGCTCGCGTC CAGGCCAGCA

   401 CGAAGGAGGC TAAGTGGGAG GGGTAGCGGT CGTTGTCCAC TAGGGGGTCC

   451 ACTCGCTCCA GGGTGTGAAG ACACATGTCG CCCTCTTCGG CATCAAGGAA

   501 GGTGATTGGT TTATAGGTGT AGGCCACGTG ACCGGGTGTT CCTGAAGGGG

   551 GGCTATAAAA GGGGGTGGGG GCGCGTTCGT CCTCACTCTC TTCCGCATCG

   601 CTGTCTGCGA GGGCCAGCTG TTGGGCTCGC GGTTGAGGAC AAACTCTTCG

   651 CGGTCTTTCC AGTACTCTTG GATCGGAAAC CCGTCGGCCT CCGAACGGTA

   701 CTCCGCCACC GAGGGACCTG AGCGAGTCCG CATCGACCGG ATCGGAAAAC

   751 CTCTCGAGAA AGGCGTCTAA CCAGTCACAG TCGCAAGGTA GGCTGAGCAC

   801 CGTGGCGGGC GGCAGCGGGT GGCGGTCGGG GTTGTTTCTG GCGGAGGTGC

   851 TGCTGATGAT GTAATTAAAG TAGGCGGTCT TGAGACGGCG GATGGTCGAG

   901 GTGAGGTGTG GCAGGCTTGA GATCGATCTG GCCATACACT TGAGTGACAA

   951 TGACATCCAC TTTGCCTTTC TCTCCACAGG TGTCCACTCC CAGGTCCAAC

  1001 TGGATCCAAG CTTCGACTCG AGGAATTCCC CGAAGGAACA AAGCACCCTC

  1051 CCCACTGGGC TCCTGGTTGC AGAGCTCCAA GTCCTCACAC AGATACGCCT

  1101 GTTTGAGAAG CAGCGGGCAA GAAAGACGCA AGCCCAGAGG CCCTGCCATT

  1151 TCTGTGGGCT CAGGTCCCTA CTGGCTCAGG CCCCTGCCTC CCTCGGCAAG

                      MET
  1201 GCCACAATGA ACCGGGGAGT CCCTTTTAGG CACTTGCTTC TGGTGCTGCA
              AA-23

  1251 ACTGGCGCTC CTCCCAGCAG CCACTCAGGG AAAGAAAGTG GTGCTGGGCA
                            AA1

  1301 AAAAAGGGGA TACAGTGGAA CTGACCTGTA CAGCTTCCCA GAAGAAGAGC

  1351 ATACAATTCC ACTGGAAAAA CTCCAACCAG ATAAAGATTC TGGGAAATCA

FIG. 16(cont'd)

```
1401  GGGCTCCTTC TTAACTAAAG GTCCATCCAA GCTGAATGAT CGCGCTGACT
1451  CAAGAAGAAG CTTGTGGGAC CAAGGAAACT TTCCCCTGAT CATCAAGAAT
1501  CTTAAGATAG AAGACTCAGA TACTTACATC TGTGAAGTGG AGGACCAGAA
1551  GGAGGAGGTG CAATTGCTAG TGTTCGGATT GACTGCCAAC TCTGACACCC
1601  ACCTGCTTCA GGGGCAGAGC CTGACCCTGA CCTTGGAGAG CCCCCCTGGT
1651  AGTAGCCCCT CAGTGCAATG TAGGAGTCCA AGGGGTAAAA ACATACAGGG
1701  GGGGAAGACC CTCTCCGTGT CTCAGCTGGA GCTCCAGGAT AGTGGCACCT
1751  GGACATGCAC TGTCTTGCAG AACCAGAAGA AGGTGGAGTT CAAAATAGAC
1801  ATCGTGGTGC TAGCTTTCCA GAACCTCCAG CATAGTCTAT AAGAAAGAGG
1851  GGGAACAGGT GGAGTTCTCC TTCCCACTCG CCTTTACAGT TGAAAAGCTG
1901  ACGGGCAGTG GCGAGCTGTG GTGGCAGGCG GAGAGGGCTT CCTCCTCCAA
1951  GTCTTGGATC ACCTTTGACC TGAAGAACAA GGAAGTGTCT GTAAAACGGG
2001  TTACCCAGGA CCCTAAGCTC CAGATGGGCA AGAAGCTCCC GCTCCACCTC
2051  ACCCTGCCCC AGGCCTTGCC TCAGTATGCT GGCTCTGGAA ACCTCACCCT
2101  GGCCCTTGAA GCGAAAACAG GAAAGTTGCA TCAGGAAGTG AACCTGGTGG
2151  TGATGAGAGC CACTCAGCTC CAGAAAAATT TGACCTGTGA GGTGTGGGGA
2201  CCCACCTCCC CTAAGCTGAT GCTGAGTTTG AAACTGGAGA ACAAGGAGGC
2251  AAAGGTCTCG AAGCGGGAGA AGGCGGTGTG GGTGCTGAAC CCTGAGGCGG
2301  GGATGTGGCA GTGTCTGCTG AGTGACTCGG GACAGGTCCT GCTGGAATCC
2351  AACATCAAGG TTCTGCCCAC ATGGTCGACC CCGGTGCAGC CAATGGCCCT
2401  GATTTGAGAT CTTTGTGAAG GAACCTTACT TCTGTGGTGT GACATAATTG
2451  GACAAACTAC CTACAGAGAT TTAAAGCTCT AAGGTAAATA TAAAATTTTT
2501  AAGTGTATAA TGTGTTAAAC TACTGATTCT AATTGTTTGT GTATTTTAGA
2551  TTCCAACCTA TGGAACTGAT GAATGGGAGC AGTGGTGGAA TGCCTTTAAT
2601  GAGGAAAACC TGTTTTGCTC AGAAGAAATG CCATCTAGTG ATGATGAGGC
2651  TACTGCTGAC TCTCAACATT CTACTCCTCC AAAAAAGAAG AGAAAGGTAG
2701  AAGACCCCAA GGACTTTCCT TCAGAATTGC TAAGTTTTTT GAGTCATGCT
2751  GTGTTTAGTA ATAGAACTCT TGCTTGCTTT GCTATTTACA CCACAAAGGA
2801  AAAAGCTGCA CTGCTATACA AGAAAATTAT GGAAAAATAT TCTGTAACCT
2851  TTATAAGTAG GCATAACAGT TATAATCATA ACATACTGTT TTTTCTTACT
2901  CCACACAGGC ATAGAGTGTC TGCTATTAAT AACTATGCTC AAAAATTGTG
2951  TACCTTTAGC TTTTTAATTT GTAAAGGGGT TAATAAGGAA TATTTGATGT
3001  ATAGTGCCTT GACTAGAGAT CATAATCAGC CATACCACAT TTGTAGAGGT
```

FIG. 16(cont'd)

```
3051  TTTACTTGCT  TTAAAAAACC  TCCCACACCT  CCCCCTGAAC  CTGAAACATA

3101  AAATGAATGC  AATTGTTGTT  GTTAACTTGT  TTATTGCAGC  TTATAATGGT

3151  TACAAATAAA  GCAATAGCAT  CACAAATTTC  ACAAATAAAG  CATTTTTTTC

3201  ACTGCATTCT  AGTTGTGGTT  TGTCCAAACT  CATCAATGTA  TCTTATCATG

3251  TCTGGATCCT  CTACGCCGGA  CGCATCGTGG  CCGGCATCAC  CGGCGCCACA

3301  GGTGCGGTTG  CTGGCGCCTA  TATCGCCGAC  ATCACCGATG  GGGAAGATCG

3351  GGCTCGCCAC  TTCGGGCTCA  TGAGCGCTTG  TTTCGGCGTG  GGTATGGTGG

3401  CAGGCCCGTG  GCCGGGGGAC  TGTTGGGCGC  CATCTCCTTG  CATGCACCAT

3451  TCCTTGCGGC  GGCGGTGCTC  AACGGCCTCA  ACCTACTACT  GGGCTGCTTC

3501  CTAATGCAGG  AGTCGCATAA  GGGAGAGCGT  CGACCGATGC  CCTTGAGAGC

3551  CTTCAACCCA  GTCAGCTCCT  TCCGGTGGGC  GCGGGGCATG  ACTATCGTCG

3601  CCGCACTTAT  GACTGTCTTC  TTTATCATGC  AACTCGTAGG  ACAGGTGCCG

3651  GCAGCGCTCT  GGGTCATTTT  CGGCGAGGAC  CGCTTTCGCT  GGAGCGCGAC

3701  GATGATCGGC  CTGTCGCTTG  CGGTATTCGG  AATCTTGCAC  GCCCTCGCTC

3751  AAGCCTTCGT  CACTGGTCCC  GCCACCAAAC  GTTTCGGCGA  GAAGCAGGCC

3801  ATTATCGCCG  GCATGGCGGC  CGACGCGCTG  GGCTACGTCT  TGCTGGCGTT

3851  CGCGACGCGA  GGCTGGATGG  CCTTCCCCAT  TATGATTCTT  CTCGCTTCCG

3901  GCGGCATCGG  GATGCCCGCG  TTGCAGGCCA  TGCTGTCCAG  GCAGGTAGAT

3951  GACGACCATC  AGGGACAGCT  TCAAGGATCG  CTCGCGGCTC  TTACCAGCCT

4001  AACTTCGATC  ACTGGACCGC  TGATCGTCAC  GGCGATTTAT  GCCGCCTCGG

4051  CGAGCACATG  GAACGGGTTG  GCATGGATTG  TAGGCGCCGC  CCTATACCTT

4101  GTCTGCCTCC  CCGCGTTGCG  TCGCGGTGCA  TGGAGCCGGG  CCACCTCGAC

4151  CTGAATGGAA  GCCGGCGGCA  CCTCGCTAAC  GGATTCACCA  CTCCAAGAAT

4201  TGGAGCCAAT  CAATTCTTGC  GGAGAACTGT  GAATGCGCAA  ACCAACCCTT

4251  GGCAGAACAT  ATCCATCGCG  TCCGCCATCT  CCAGCAGCCG  CACGCGGCGC

4301  ATCTCGGGCC  GCGTTGCTGG  CGTTTTTCCA  TAGGCTCCGC  CCCCCTGACG

4351  AGCATCACAA  AAATCGACGC  TCAAGTCAGA  GGTGGCGAAA  CCCGACAGGA

4401  CTATAAAGAT  ACCAGGCGTT  TCCCCCTGGA  AGCTCCCTCG  TGCGCTCTCC

4451  TGTTCCGACC  CTGCCGCTTA  CCGGATACCT  GTCCGCCTTT  CTCCCTTCGG

4501  GAAGCGTGGC  GCTTTCTCAA  TGCTCACGCT  GTAGGTATCT  CAGTTCGGTG

4551  TAGGTCGTTC  GCTCCAAGCT  GGGCTGTGTG  CACGAACCCC  CCGTTCAGCC

4601  CGACCGCTGC  GCCTTATCCG  GTAACTATCG  TCTTGAGTCC  AACCCGGTAA

4651  GACACGACTT  ATCGCCACTG  GCAGCAGCCA  CTGGTAACAG  GATTAGCAGA
```

FIG. 16(cont'd)

```
4701  GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA

4751  CGGCTACACT AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG

4801  TTACCTTCGG AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACCACC

4851  GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG CAGCAGATTA CGCGCAGAAA

4901  AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC

4951  AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA

5001  AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT

5051  CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA

5101  GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC

5151  TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG

5201  CCCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT

5251  TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT

5301  GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAGCTAG

5351  AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGTT GCCATTGCTG

5401  CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC

5451  GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT TGTGCAAAAA

5501  AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG

5551  CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC

5601  ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC

5651  ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC CCGGCGTCAA

5701  CACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT

5751  GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG

5801  ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT

5851  TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC

5901  GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT

5951  CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG

6001  GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC

6051  ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA TTATTATCAT

6101  GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGTCTTCAA
```

FIG. 17
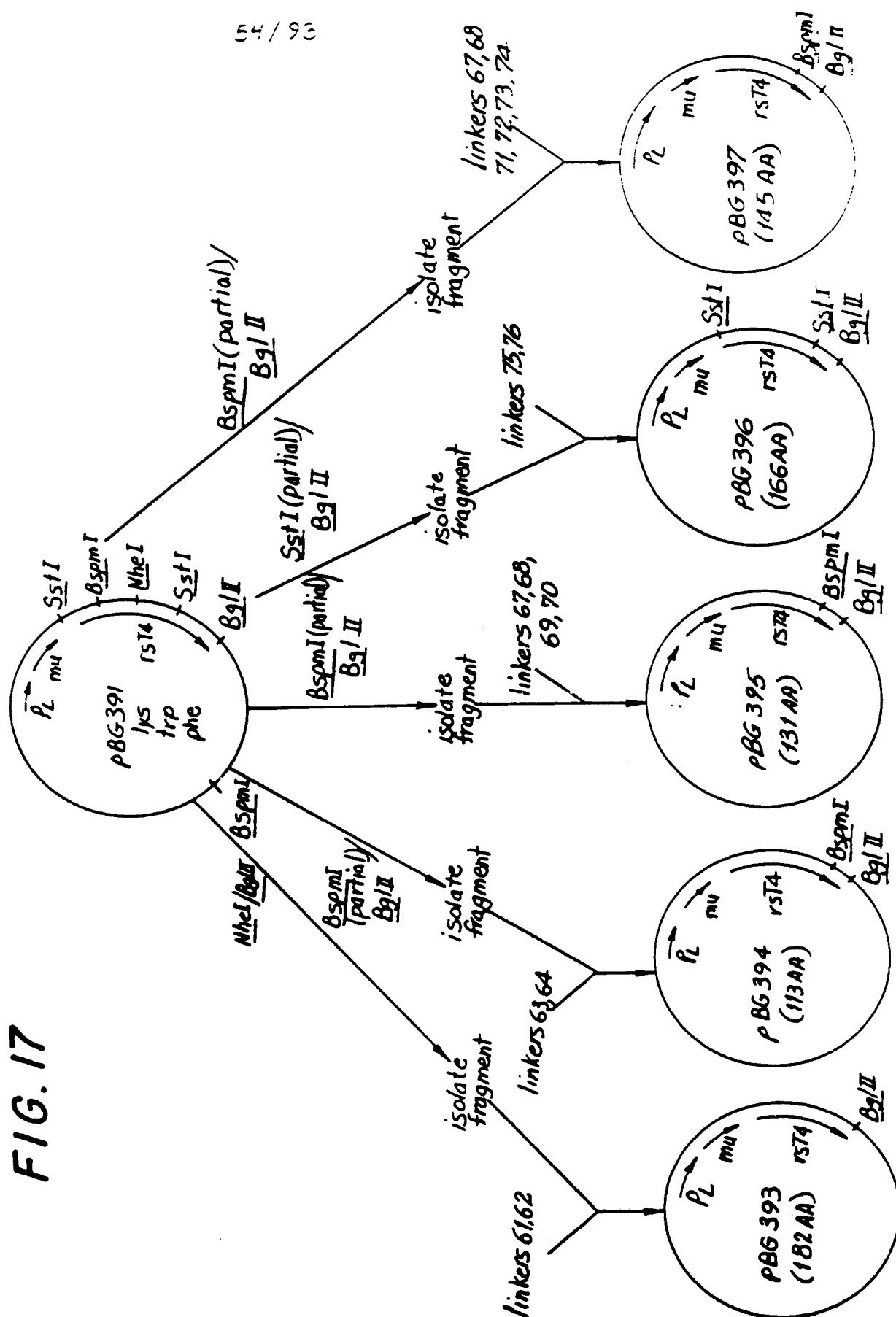
FIG.18

61 5' CTA CCT TTT CCA CTC A 3'

62 5' GAT CTC ACT GGA AAG 3'

63 5' CCG CTG ATA CTA A 3'

64 5' GAT CTT ACT ATC A 3'

67 5' CGG CCA GAG CCT GAC CCT GAC CTT GGA GAG CCC C 3'

68 5' CCG GGG GGC TCT CCA AGG TCA GGG TCA GGC TCT G 3'

69 5' CCG GGT AGT AGC CCC TCA GTG CAA TGA 3'

70 5' GAT CTC ATT GCA CTG AGG GGC TAC TAC 3'

71 5' CCG GGT AGT AGC CCC TCA GTG CAA TGT AGG AGT C 3'

72 5' TAG GAC TCC TAC ATT GCA CTG AGG GGC TAC TAC 3'

73 5' CTA GGG GTA AAA ACA TAC AGG GGG GGA AGA CCT GA 3'

74 5' GAT CTC AGG TCT TTC CCC CCC TGT ATG TTT TTA CCC 3'

75 5' CCA GGA TAG TGG CAC CTC GAC ATG CAC TGT CTT CCA CAA CTG A 3'

76 5' GAT CTC AGT TGT GGA AGA CAG TGC ATG TCG AGG TGC CAC TAT CCT GGA CCT 3'

pBG394
:BG368 backbone
:soluble T4S
:AA #3 = LVS
:first 113 AA of T4
:basically LD to VIJI

FIG. 19

bg394.seq   Length: 5365

```
  1  GAATTAATTC CAGCTTGCTG TGGAATGTGT GTCAGTTAGG GTGTGGAAAG

 51  TCCCCAGGCT CCCCAGCAGG CAGAAGTATG CAAAGCATGC ATCTCAATTA

101  GTCAGCAACC AGGTGTGGAA AGTCCCCAGG CTCCCCAGCA GGCAGAAGTA

151  TGCAAAGCAT GCATCTCAAT TAGTCAGCAA CCATAGTCCC GCCCCTAACT

201  CCGCCCATCC CGCCCCTAAC TCCGCCCAGT TCCGCCCATT CTCCGCCCCA

251  TGGCTGACTA ATTTTTTTTA TTTATGCAGA GGCCGAGGCC GCCTCGGCCT

301  CTGAGCTATT CCAGAAGTAG TGAGGAGGCT TTTTTGGAGG GGTCCTCCTC

351  GTATAGAAAC TCGGACCACT CTGAGACGAA GGCTCGCGTC CAGGCCAGCA

401  CGAAGGAGGC TAAGTGGGAG GGGTAGCGGT CGTTGTCCAC TAGGGGGTCC

451  ACTCGCTCCA GGGTGTGAAG ACACATGTCG CCCTCTTCGG CATCAAGGAA

501  GGTGATTGGT TTATAGGTGT AGGCCACGTG ACCGGGTGTT CCTGAAGGGG

551  GGCTATAAAA GGGGGTGGGG GCGCGTTCGT CCTCACTCTC TTCCGCATCG

601  CTGTCTGCGA GGGCCAGCTG TTGGGCTCGC GGTTGAGGAC AAACTCTTCG

651  CGGTCTTTCC AGTACTCTTG GATCGGAAAC CCGTCGGCCT CCGAACGGTA

701  CTCCGCCACC GAGGGACCTG AGCGAGTCCG CATCGACCGG ATCGGAAAAC

751  CTCTCGAGAA AGGCGTCTAA CCAGTCACAG TCGCAAGGTA GGCTGAGCAC

801  CGTGGCGGGC GGCAGCGGGT GGCGGTCGGG GTTGTTTCTG GCGGAGGTGC

851  TGCTGATGAT GTAATTAAAG TAGGCGGTCT TGAGACGGCG GATGGTCGAG

901  GTGAGGTGTG GCAGGCTTGA GATCGATCTG GCCATACACT TGAGTGACAA

951  TGACATCCAC TTTGCCTTTC TCTCCACAGG TGTCCACTCC CAGGTCCAAC

1001  TGGATCCAAG CTTCGACTCG AGGAATTCCC CGAAGGAACA AAGCACCCTC

1051  CCCACTGGGC TCCTGGTTGC AGAGCTCCAA GTCCTCACAC AGATACGCCT

1101  GTTTGAGAAG CAGCGGGCAA GAAAGACGCA AGCCCAGAGG CCCTGCCATT

1151  TCTGTGGGCT CAGGTCCCTA CTGGCTCAGG CCCCTGCCTC CCTCGGCAAG

1201  GCCACAATGA ACCGGGGAGT CCCTTTTAGG CACTTGCTTC TGGTGCTGCA

1251  ACTGGCGCTC CTCCCAGCAG CCACTCAGGG AAAGAAAGTG GTGCTGGGCA

1301  AAAAAGGGGA TACAGTGGAA CTGACCTGTA CAGCTTCCCA GAAGAAGAGC

1351  ATACAATTCC ACTGGAAAAA CTCCAACCAG ATAAAGATTC TGGGAAATCA
```

FIG. 19 (cont'd)

```
1401  GGGCTCCTTC TTAACTAAAG GTCCATCCAA GCTGAATGAT CGCGCTGACT

1451  CAAGAAGAAG CTTGTGGGAC CAAGGAAACT TTCCCCTGAT CATCAAGAAT

1501  CTTAAGATAG AAGACTCAGA TACTTACATC TGTGAAGTGG AGGACCAGAA

1551  GGAGGAGGTG CAATTGCTAG TGTTCGGATT GACTGCCAAC TCTGACACCC

1601  ACCTGCTTCA GGGGTGATAG TAAGATCTTT GTGAAGGAAC CTTACTTCTG

1651  TGGTGTGACA TAATTGGACA AACTACCTAC AGAGATTTAA AGCTCTAAGG

1701  TAAATATAAA ATTTTTAAGT GTATAATGTG TTAAACTACT GATTCTAATT

1751  GTTTGTGTAT TTTAGATTCC AACCTATGGA ACTGATGAAT GGGAGCAGTG

1801  GTGGAATGCC TTTAATGAGG AAAACCTGTT TTGCTCAGAA GAAATGCCAT

1851  CTAGTGATGA TGAGGCTACT GCTGACTCTC AACATTCTAC TCCTCCAAAA

1901  AAGAAGAGAA AGGTAGAAGA CCCCAAGGAC TTTCCTTCAG AATTGCTAAG

1951  TTTTTTGAGT CATGCTGTGT TTAGTAATAG AACTCTTGCT TGCTTTGCTA

2001  TTTACACCAC AAAGGAAAAA GCTGCACTGC TATACAAGAA AATTATGGAA

2051  AAATATTCTG TAACCTTTAT AAGTAGGCAT AACAGTTATA ATCATAACAT

2101  ACTGTTTTTT CTTACTCCAC ACAGGCATAG AGTGTCTGCT ATTAATAACT

2151  ATGCTCAAAA ATTGTGTACC TTTAGCTTTT TAATTTGTAA AGGGGTTAAT

2201  AAGGAATATT TGATGTATAG TGCCTTGACT AGAGATCATA ATCAGCCATA

2251  CCACATTTGT AGAGGTTTTA CTTGCTTTAA AAAACCTCCC ACACCTCCCC

2301  CTGAACCTGA AACATAAAAT GAATGCAATT GTTGTTGTTA ACTTGTTTAT

2351  TGCAGCTTAT AATGGTTACA AATAAAGCAA TAGCATCACA AATTTCACAA

2401  ATAAAGCATT TTTTTCACTG CATTCTAGTT GTGGTTTGTC CAAACTCATC

2451  AATGTATCTT ATCATGTCTG GATCCTCTAC GCCGGACGCA TCGTGGCCGG

2501  CATCACCGGC GCCACAGGTG CGGTTGCTGG CGCCTATATC GCCGACATCA

2551  CCGATGGGGA AGATCGGGCT CGCCACTTCG GGCTCATGAG CGCTTGTTTC

2601  GGCGTGGGTA TGGTGGCAGG CCCGTGGCCG GGGGACTGTT GGGCGCCATC

2651  TCCTTGCATG CACCATTCCT TGCGGCGGCG GTGCTCAACG GCCTCAACCT

2701  ACTACTGGGC TGCTTCCTAA TGCAGGAGTC GCATAAGGGA GAGCGTCGAC

2751  CGATGCCCTT GAGAGCCTTC AACCCAGTCA GCTCCTTCCG GTGGGCGCGG

2801  GGCATGACTA TCGTCGCCGC ACTTATGACT GTCTTCTTTA TCATGCAACT

2851  CGTAGGACAG GTGCCGGCAG CGCTCTGGGT CATTTTCGGC GAGGACCGCT

2901  TTCGCTGGAG CGCGACGATG ATCGGCCTGT CGCTTGCGGT ATTCGGAATC

2951  TTGCACGCCC TCGCTCAAGC CTTCGTCACT GGTCCCGCCA CCAAACGTTT

3001  CGGCGAGAAG CAGGCCATTA TCGCCGGCAT GGCGGCCGAC GCGCTGGGCT
```

FIG.19 (cont'd)

```
305:  ACGTCTTGCT GGCGTTCGCG ACGCGAGGCT GGATGGCCTT CCCCATTATG
310:  ATTCTTCTCG CTTCCGGCGG CATCGGGATG CCCGCGTTGC AGGCCATGCT
315:  GTCCAGGCAG GTAGATGACG ACCATCAGGG ACAGCTTCAA GGATCGCTCG
320:  CGGCTCTTAC CAGCCTAACT TCGATCACTG GACCGCTGAT CGTCACGGCG
325:  ATTTATGCCG CCTCGGCGAG CACATGGAAC GGGTTGGCAT GGATTGTAGG
330:  CGCCGCCCTA TACCTTGTCT GCCTCCCCGC GTTGCGTCGC GGTGCATGGA
3351: GCCGGGCCAC CTCGACCTGA ATGGAAGCCG GCGGCACCTC GCTAACGGAT
3401: TCACCACTCC AAGAATTGGA GCCAATCAAT TCTTGCGGAG AACTGTGAAT
3451: GCGCAAACCA ACCCTTGGCA GAACATATCC ATCGCGTCCG CCATCTCCAG
3501: CAGCCGCACG CGGCGCATCT CGGGCCGCGT TGCTGGCGTT TTTCCATAGG
3551: CTCCGCCCCC CTGACGAGCA TCACAAAAAT CGACGCTCAA GTCAGAGGTG
3601: GCGAAACCCG ACAGGACTAT AAAGATACCA GGCGTTTCCC CCTGGAAGCT
3651: CCCTCGTGCG CTCTCCTGTT CCGACCCTGC CGCTTACCGG ATACCTGTCC
3701: GCCTTTCTCC CTTCGGGAAG CGTGGCGCTT TCTCAATGCT CACGCTGTAG
3751: GTATCTCAGT TCGGTGTAGG TCGTTCGCTC CAAGCTGGGC TGTGTGCACG
3801: AACCCCCCGT TCAGCCCGAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT
3851: GAGTCCAACC CGGTAAGACA CGACTTATCG CCACTGGCAG CAGCCACTGG
3901: TAACAGGATT AGCAGAGCGA GGTATGTAGG CGGTGCTACA GAGTTCTTGA
3951: AGTGGTGGCC TAACTACGGC TACACTAGAA GGACAGTATT TGGTATCTGC
4001: GCTCTGCTGA AGCCAGTTAC CTTCGGAAAA AGAGTTGGTA GCTCTTGATC
4051: CGGCAAACAA ACCACCGCTG GTAGCGGTGG TTTTTTTGTT TGCAAGCAGC
4101: AGATTACGCG CAGAAAAAAA GGATCTCAAG AAGATCCTTT GATCTTTTCT
4151: ACGGGGTCTG ACGCTCAGTG GAACGAAAAC TCACGTTAAG GGATTTTGGT
4201: CATGAGATTA TCAAAAGGA TCTTCACCTA GATCCTTTTA AATTAAAAAT
4251: GAAGTTTTAA ATCAATCTAA AGTATATATG AGTAAACTTG GTCTGACAGT
4301: TACCAATGCT TAATCAGTCA GGCACCTATC TCAGCGATCT GTCTATTTCG
4351: TTCATCCATA GTTGCCTGAC TCCCCGTCGT GTAGATAACT ACGATACGGG
4401: AGGGCTTACC ATCTGGCCCC AGTGCTGCAA TGATACCGCG AGACCCACGC
4451: TCACCGGCTC CAGATTTATC AGCAATAAAC CAGCCAGCCG GAAGGGCCGA
4501: GCGCAGAAGT GGTCCTGCAA CTTTATCCGC CTCCATCCAG TCTATTAATT
4551: GTTGCCGGGA AGCTAGAGTA AGTAGTTCGC CAGTTAATAG TTTGCGCAAC
4601: GTTGTTGCCA TTGCTGCAGG CATCGTGGTG TCACGCTCGT CGTTTGGTAT
4651: GGCTTCATTC AGCTCCGGTT CCCAACGATC AAGGCGAGTT ACATGATCCC
```

FIG. 19 (cont'd)

```
4701  CCATGTTGTG  CAAAAAAGCG  GTTAGCTCCT  TCGGTCCTCC  GATCGTTGTC

4751  AGAAGTAAGT  TGGCCGCAGT  GTTATCACTC  ATGGTTATGG  CAGCACTGCA

4801  TAATTCTCTT  ACTGTCATGC  CATCCGTAAG  ATGCTTTTCT  GTGACTGGTG

4851  AGTACTCAAC  CAAGTCATTC  TGAGAATAGT  GTATGCGGCG  ACCGAGTTGC

4901  TCTTGCCCGG  CGTCAACACG  GGATAATACC  GCGCCACATA  GCAGAACTTT

4951  AAAAGTGCTC  ATCATTGGAA  AACGTTCTTC  GGGGCGAAAA  CTCTCAAGGA

5001  TCTTACCGCT  GTTGAGATCC  AGTTCGATGT  AACCCACTCG  TGCACCCAAC

5051  TGATCTTCAG  CATCTTTTAC  TTTCACCAGC  GTTTCTGGGT  GAGCAAAAAC

5101  AGGAAGGCAA  AATGCCGCAA  AAAAGGGAAT  AAGGGCGACA  CGGAAATGTT

5151  GAATACTCAT  ACTCTTCCTT  TTTCAATATT  ATTGAAGCAT  TTATCAGGGT

5201  TATTGTCTCA  TGAGCGGATA  CATATTTGAA  TGTATTTAGA  AAAATAAACA

5251  AATAGGGGTT  CCGCGCACAT  TTCCCCGAAA  AGTGCCACCT  GACGTCTAAG

5301  AAACCATTAT  TATCATGACA  TTAACCTATA  AAAATAGGCG  TATCACGAGG

5351  CCCTTTCGTC  TTCAA
```

SUBSTITUTE SHEET

pBG396
:BG368 backbon
:soluble T4#12
:AA #3 = LYS

FIG. 20

```
bg396.seq  Length: 5518

   1  GAATTAATTC CAGCTTGCTG TGGAATGTGT GTCAGTTAGG GTGTGGAAAG
  51  TCCCCAGGCT CCCCAGCAGG CAGAAGTATG CAAAGCATGC ATCTCAATTA
 101  GTCAGCAACC AGGTGTGGAA AGTCCCCAGG CTCCCCAGCA GGCAGAAGTA
 151  TGCAAAGCAT GCATCTCAAT TAGTCAGCAA CCATAGTCCC GCCCCTAACT
 201  CCGCCCATCC CGCCCCTAAC TCCGCCCAGT TCCGCCCATT CTCCGCCCCA
 251  TGGCTGACTA ATTTTTTTTA TTTATGCAGA GGCCGAGGCC GCCTCGGCCT
 301  CTGAGCTATT CCAGAAGTAG TGAGGAGGCT TTTTTGGAGG GGTCCTCCTC
 351  GTATAGAAAC TCGGACCACT CTGAGACGAA GGCTCGCGTC CAGGCCAGCA
 401  CGAAGGAGGC TAAGTGGGAG GGGTAGCGGT CGTTGTCCAC TAGGGGGTCC
 451  ACTCGCTCCA GGGTGTGAAG ACACATGTCG CCCTCTTCGG CATCAAGGAA
 501  GGTGATTGGT TTATAGGTGT AGGCCACGTG ACCGGGTGTT CCTGAAGGGG
 551  GGCTATAAAA GGGGGTGGGG CGCGTTCGT CCTCACTCTC TTCCGCATCG
 601  CTGTCTGCGA GGGCCAGCTG TTGGGCTCGC GGTTGAGGAC AAACTCTTCG
 651  CGGTCTTTCC AGTACTCTTG GATCGGAAAC CCGTCGGCCT CCGAACGGTA
 701  CTCCGCCACC GAGGGACCTG AGCGAGTCCG CATCGACCGG ATCGGAAAAC
 751  CTCTCGAGAA AGGCGTCTAA CCAGTCACAG TCGCAAGGTA GGCTGAGCAC
 801  CGTGGCGGGC GGCAGCGGGT GGCGGTCGGG GTTGTTTCTG GCGGAGGTGC
 851  TGCTGATGAT GTAATTAAAG TAGGCGGTCT TGAGACGGCG GATGGTCGAG
 901  GTGAGGTGTG GCAGGCTTGA GATCGATCTG GCCATACACT TGAGTGACAA
 951  TGACATCCAC TTTGCCTTTC TCTCCACAGG TGTCCACTCC CAGGTCCAAC
1001  TGGATCCAAG CTTCGACTCG AGGAATTCCC CGAAGGAACA AAGCACCCTC
1051  CCCACTGGGC TCCTGGTTGC AGAGCTCCAA GTCCTCACAC AGATACGCCT
1101  GTTTGAGAAG CAGCGGGCAA GAAAGACGCA AGCCCAGAGG CCCTGCCATT
1151  TCTGTGGGCT CAGGTCCCTA CTGGCTCAGG CCCCTGCCTC CCTCGGCAAG
1201  GCCACAATGA ACCGGGGAGT CCCTTTTAGG CACTTGCTTC TGGTGCTGCA
1251  ACTGGCGCTC CTCCCAGCAG CCACTCAGGG AAAGAAAGTG GTGCTGGGCA
1301  AAAAAGGGGA TACAGTGGAA CTGACCTGTA CAGCTTCCCA GAAGAAGAGC
1351  ATACAATTCC ACTGGAAAAA CTCCAACCAG ATAAAGATTC TGGGAAATCA
1401  GGGCTCCTTC TTAACTAAAG GTCCATCCAA GCTGAATGAT CGCGCTGACT
```

SUBSTITUTE SHEET

FIG. 20 (cont'd)

```
1451  CAAGAAGAAG CTTGTGGGAC CAAGGAAACT TTCCCCTGAT CATCAAGAAT

1501  CTTAAGATAG AAGACTCAGA TACTTACATC TGTGAAGTGG AGGACCAGAA

1551  GGAGGAGGTG CAATTGCTAG TGTTCGGATT GACTGCCAAC TCTGACACCC

1601  ACCTGCTTCA GGGGCAGAGC CTGACCCTGA CCTTGGAGAG CCCCCCTGGT

1651  AGTAGCCCCT CAGTGCAATG TAGGAGTCCA AGGGGTAAAA ACATACAGGG

1701  GGGGAAGACC CTCTCCCTGT CTCAGCTGGA GCTCCAGGAT AGTGGCACCT

1751  GGACATGCAC TGTCTTGCAG AACTGAGATC TTTGTGAAGG AACCTTACTT

1801  CTGTGGTGTG ACATAATTGG ACAAACTACC TACAGAGATT TAAAGCTCTA

1851  AGGTAAATAT AAAATTTTTA AGTGTATAAT GTGTTAAACT ACTGATTCTA

1901  ATTGTTTGTG TATTTTAGAT TCCAACCTAT GGAACTGATG AATGGGAGCA

1951  GTGGTGGAAT GCCTTTAATG AGGAAAACCT GTTTTGCTCA GAAGAAATGC

2001  CATCTAGTGA TGATGAGGCT ACTGCTGACT CTCAACATTC TACTCCTCCA

2051  AAAAAGAAGA GAAAGGTAGA AGACCCCAAG GACTTTCCTT CAGAATTGCT

2101  AAGTTTTTTG AGTCATGCTG TGTTTAGTAA TAGAACTCTT GCTTGCTTTG

2151  CTATTTACAC CACAAAGGAA AAAGCTGCAC TGCTATACAA GAAAATTATG

2201  GAAAAATATT CTGTAACCTT TATAAGTAGG CATAACAGTT ATAATCATAA

2251  CATACTGTTT TTTCTTACTC CACACAGGCA TAGAGTGTCT GCTATTAATA

2301  ACTATGCTCA AAAATTGTGT ACCTTTAGCT TTTTAATTTG TAAAGGGGTT

2351  AATAAGGAAT ATTTGATGTA TAGTGCCTTG ACTAGAGATC ATAATCAGCC

2401  ATACCACATT TGTAGAGGTT TTACTTGCTT TAAAAAACCT CCCACACCTC

2451  CCCCTGAACC TGAAACATAA AATGAATGCA ATTGTTGTTG TTAACTTGTT

2501  TATTGCAGCT TATAATGGTT ACAAATAAAG CAATAGCATC ACAAATTTCA

2551  CAAATAAAGC ATTTTTTTCA CTGCATTCTA GTTGTGGTTT GTCCAAACTC

2601  ATCAATGTAT CTTATCATGT CTGGATCCTC TACGCCGGAC GCATCGTGGC

2651  CGGCATCACC GGCGCCACAG GTGCGGTTGC TGGCGCCTAT ATCGCCGACA

2701  TCACCGATGG GGAAGATCGG GCTCGCCACT TCGGGCTCAT GAGCGCTTGT

2751  TTCGGCGTGG GTATGGTGGC AGGCCCGTGG CCGGGGGACT GTTGGGCGCC

2801  ATCTCCTTGC ATGCACCATT CCTTGCGGCG GCGGTGCTCA ACGGCCTCAA

2851  CCTACTACTG GGCTGCTTCC TAATGCAGGA GTCGCATAAG GGAGAGCGTC

2901  GACCGATGCC CTTGAGAGCC TTCAACCCAG TCAGCTCCTT CCGGTGGGCG

2951  CGGGGCATGA CTATCGTCGC CGCACTTATG ACTGTCTTCT TTATCATGCA

3001  ACTCGTAGGA CAGGTGCCGG CAGCGCTCTG GGTCATTTTC GGCGAGGACC

3051  GCTTTCGCTG GAGCGCGACG ATGATCGGCC TGTCGCTTGC GGTATTCGGA
```

```
3101   ATCTTGCACG  CCCTCGCTCA  AGCCTTCGTC  ACTGGTCCCG  CCACCAAACG

3151   TTTCGGCGAG  AAGCAGGCCA  TTATCGCCGG  CATGGCGGCC  GACGCGCTGG

3201   GCTACGTCTT  GCTGGCGTTC  GCGACGCGAG  GCTGGATGGC  CTTCCCCATT

3251   ATGATTCTTC  TCGCTTCCGG  CGGCATCGGG  ATGCCCGCGT  TGCAGGCCAT

3301   GCTGTCCAGG  CAGGTAGATG  ACGACCATCA  GGGACAGCTT  CAAGGATCGC

3351   TCGCGGCTCT  TACCAGCCTA  ACTTCGATCA  CTGGACCGCT  GATCGTCACG

3401   GCGATTTATG  CCGCCTCGGC  GAGCACATGG  AACGGGTTGG  CATGGATTGT

3451   AGGCGCCGCC  CTATACCTTG  TCTGCCTCCC  CGCGTTGCGT  CGCGGTGCAT

3501   GCAGCCGGGC  CACCTCGACC  TGAATGGAAG  CCGGCGGCAC  CTCGCTAACG

3551   GATTCACCAC  TCCAAGAATT  GGAGCCAATC  AATTCTTGCG  GAGAACTGTG

3601   AATGCGCAAA  CCAACCCTTG  GCAGAACATA  TCCATCGCGT  CCGCCATCTC

3651   CAGCAGCCGC  ACGCGGCGCA  TCTCGGGCCG  CGTTGCTGGC  GTTTTTCCAT

3701   AGGCTCCGCC  CCCCTGACGA  GCATCACAAA  AATCGACGCT  CAAGTCAGAG

3751   GTGGCGAAAC  CCGACAGGAC  TATAAAGATA  CCAGGCGTTT  CCCCCTGGAA

3801   GCTCCCTCGT  GCGCTCTCCT  GTTCCGACCC  TGCCGCTTAC  CGGATACCTG

3851   TCCGCCTTTC  TCCCTTCGGG  AAGCGTGGCG  CTTTCTCAAT  GCTCACGCTG

3901   TAGGTATCTC  AGTTCGGTGT  AGGTCGTTCG  CTCCAAGCTG  GGCTGTGTGC

3951   ACGAACCCCC  CGTTCAGCCC  GACCGCTGCG  CCTTATCCGG  TAACTATCGT

4001   CTTGAGTCCA  ACCCGGTAAG  ACACGACTTA  TCGCCACTGG  CAGCAGCCAC

4051   TGGTAACAGG  ATTAGCAGAG  CGAGGTATGT  AGGCGGTGCT  ACAGAGTTCT

4101   TGAAGTGGTG  GCCTAACTAC  GGCTACACTA  GAAGGACAGT  ATTTGGTATC

4151   TGCGCTCTGC  TGAAGCCAGT  TACCTTCGGA  AAAAGAGTTG  GTAGCTCTTG

4201   ATCCGGCAAA  CAAACCACCG  CTGGTAGCGG  TGGTTTTTTT  GTTTGCAAGC

4251   AGCAGATTAC  GCGCAGAAAA  AAAGGATCTC  AAGAAGATCC  TTTGATCTTT

4301   TCTACGGGGT  CTGACGCTCA  GTGGAACGAA  AACTCACGTT  AAGGGATTTT

4351   GGTCATGAGA  TTATCAAAAA  GGATCTTCAC  CTAGATCCTT  TTAAATTAAA

4401   AATGAAGTTT  TAAATCAATC  TAAAGTATAT  ATGAGTAAAC  TTGGTCTGAC

4451   AGTTACCAAT  GCTTAATCAG  TGAGGCACCT  ATCTCAGCGA  TCTGTCTATT

4501   TCGTTCATCC  ATAGTTGCCT  GACTCCCCGT  CGTGTAGATA  ACTACGATAC

4551   GGGAGGGCTT  ACCATCTGGC  CCCAGTGCTG  CAATGATACC  GCGAGACCCA

4601   CGCTCACCGG  CTCCAGATTT  ATCAGCAATA  AACCAGCCAG  CCGGAAGGGC

4651   CGAGCGCAGA  AGTGGTCCTG  CAACTTTATC  CGCCTCCATC  CAGTCTATTA

4701   ATTGTTGCCG  GGAAGCTAGA  GTAAGTAGTT  CGCCAGTTAA  TAGTTTGCGC
```

FIG. 20 (cont'd)

```
4751  AACGTTGTTG  CCATTGCTGC  AGGCATCGTG  GTGTCACGCT  CGTCGTTTGG
4801  TATGGCTTCA  TTCAGCTCCG  GTTCCCAACG  ATCAAGGCGA  GTTACATGAT
4851  CCCCCATGTT  GTGCAAAAAA  GCGGTTAGCT  CCTTCGGTCC  TCCGATCGTT
4901  GTCAGAAGTA  AGTTGGCCGC  AGTGTTATCA  CTCATGGTTA  TGGCAGCACT
4951  .GCATAATTCT  CTTACTGTCA  TGCCATCCGT  AAGATGCTTT  TCTGTGACTG
5001  GTGAGTACTC  AACCAAGTCA  TTCTGAGAAT  AGTGTATGCG  GCGACCGAGT
5051  TGCTCTTGCC  CGGCGTCAAC  ACGGGATAAT  ACCGCGCCAC  ATAGCAGAAC
5101  TTTAAAAGTG  CTCATCATTG  GAAAACGTTC  TTCGGGGCGA  AAACTCTCAA
5151  GGATCTTACC  GCTGTTGAGA  TCCAGTTCGA  TGTAACCCAC  TCGTGCACCC
5201  AACTGATCTT  CAGCATCTTT  TACTTTCACC  AGCGTTTCTG  GGTGAGCAAA
5251  AACAGGAAGG  CAAAATGCCG  CAAAAAAGGG  AATAAGGGCG  ACACGGAAAT
5301  GTTGAATACT  CATACTCTTC  CTTTTTCAAT  ATTATTGAAG  CATTTATCAG
5351  GGTTATTGTC  TCATGAGCGG  ATACATATTT  GAATGTATTT  AGAAAAATAA
5401  ACAAATAGGG  GTTCCGCGCA  CATTTCCCCG  AAAAGTGCCA  CCTGACGTCT
5451  AAGAAACCAT  TATTATCATG  ACATTAACCT  ATAAAAATAG  GCGTATCACG
5501  AGGCCCTTTC  GTCTTCAA
```

PBG393
:BG368 backbone
:soluble T4#8
:AA #3 = LYS
:"perfect" Stu/first 182 AA of T4
:basically up to V2J2

FIG.21

bg393.seq   Length: 5566

```
   1  GAATTAATTC CAGCTTGCTG TGGAATGTGT GTCAGTTAGG GTGTGGAAAG

  51  TCCCCAGGCT CCCCAGCAGG CAGAAGTATG CAAAGCATGC ATCTCAATTA

 101  GTCAGCAACC AGGTGTGGAA AGTCCCCAGG CTCCCCAGCA GGCAGAAGTA

 151  TGCAAAGCAT GCATCTCAAT TAGTCAGCAA CCATAGTCCC GCCCCTAACT

 201  CCGCCCATCC CGCCCCTAAC TCCGCCCAGT TCCGCCCATT CTCCGCCCCA

 251  TGGCTGACTA ATTTTTTTTA TTTATGCAGA GGCCGAGGCC GCCTCGGCCT

 301  CTGAGCTATT CCAGAAGTAG TGAGGAGGCT TTTTTGGAGG GGTCCTCCTC

 351  GTATAGAAAC TCGGACCACT CTGAGACGAA GGCTCGCGTC CAGGCCAGCA

 401  CGAAGGAGGC TAAGTGGGAG GGGTAGCGGT CGTTGTCCAC TAGGGGGTCC

 451  ACTCGCTCCA GGGTGTGAAG ACACATGTCG CCCTCTTCGG CATCAAGGAA

 501  GGTGATTGGT TTATAGGTGT AGGCCACGTG ACCGGGTGTT CCTGAAGGGG

 551  GGCTATAAAA GGGGGTGGGG GCGCGTTCGT CCTCACTCTC TTCCGCATCG

 601  CTGTCTGCGA GGGCCAGCTG TTGGGCTCGC GGTTGAGGAC AAACTCTTCG

 651  CGGTCTTTCC AGTACTCTTG GATCGGAAAC CCGTCGGCCT CCGAACGGTA

 701  CTCCGCCACC GAGGGACCTG AGCGAGTCCG CATCGACCGG ATCGGAAAAC

 751  CTCTCGAGAA AGGCGTCTAA CCAGTCACAG TCGCAAGGTA GGCTGAGCAC

 801  CGTGGCGGGC GGCAGCGGGT GGCGGTCGGG GTTGTTTCTG GCGGAGGTGC

 851  TGCTGATGAT GTAATTAAAG TAGGCGGTCT TGAGACGGCG GATGGTCGAG

 901  GTGAGGTGTG GCAGGCTTGA GATCGATCTG GCCATACACT TGAGTGACAA

 951  TGACATCCAC TTTGCCTTTC TCTCCACAGG TGTCCACTCC CAGGTCCAAC

1001  TGGATCCAAG CTTCGACTCG AGGAATTCCC CGAAGGAACA AAGCACCCTC

1051  CCCACTGGGC TCCTGGTTGC AGAGCTCCAA GTCCTCACAC AGATACGCCT

1101  GTTTGAGAAG CAGCGGGCAA GAAAGACGCA AGCCCAGAGG CCCTGCCATT

1151  TCTGTGGGCT CAGGTCCCTA CTGGCTCAGG CCCCTGCCTC CCTCGGCAAG

1201  GCCACAATGA ACCGGGGAGT CCCTTTTAGG CACTTGCTTC TGGTGCTGCA

1251  ACTGGCGCTC CTCCCAGCAG CCACTCAGGG AAAGAAAGTG GTGCTGGGCA

1301  AAAAAGGGGA TACAGTGGAA CTGACCTGTA CAGCTTCCCA GAAGAAGAGC

1351  ATACAATTCC ACTGGAAAAA CTCCAACCAG ATAAAGATTC TGGGAAATCA
```

FIG. 21 (cont'd)

```
1401  GGGCTCCTTC TTAACTAAAG GTCCATCCAA GCTGAATGAT CGCGCTGACT

1451  CAAGAAGAAG CTTGTGGGAC CAAGGAAACT TTCCCCTGAT CATCAAGAAT

1501  CTTAAGATAG AAGACTCAGA TACTTACATC TGTGAAGTGG AGGACCAGAA

1551  GGAGGAGGTG CAATTGCTAG TGTTCGGATT GACTGCCAAC TCTGACACCC

1601  ACCTGCTTCA GGGGCAGAGC CTGACCCTGA CCTTGGAGAG CCCCCCTGGT

1651  AGTAGCCCCT CAGTGCAATG TAGGAGTCCA AGGGGTAAAA ACATACAGGG

1701  GGGGAAGACC CTCTCCGTGT CTCAGCTGGA GCTCCAGGAT AGTGGCACCT

1751  GGACATGCAC TGTCTTGCAG AACCAGAAGA AGGTGGAGTT CAAAATAGAC

1801  ATCGTGGTGC TAGCTTTCCA GTGAGATCTT TGTGAAGGAA CCTTACTTCT

1851  GTGGTGTGAC ATAATTGGAC AAACTACCTA CAGAGATTTA AAGCTCTAAG

1901  GTAAATATAA AATTTTTAAG TGTATAATGT GTTAAACTAC TGATTCTAAT

1951  TGTTTGTGTA TTTTAGATTC CAACCTATGG AACTGATGAA TGGGAGCAGT

2001  GGTGGAATGC CTTTAATGAG GAAAACCTGT TTTGCTCAGA AGAAATGCCA

2051  TCTAGTGATG ATGAGGCTAC TGCTGACTCT CAACATTCTA CTCCTCCAAA

2101  AAAGAAGAGA AAGGTAGAAG ACCCCAAGGA CTTTCCTTCA GAATTGCTAA

2151  GTTTTTTGAG TCATGCTGTG TTTAGTAATA GAACTCTTGC TTGCTTTGCT

2201  ATTTACACCA CAAAGGAAAA AGCTGCACTG CTATACAAGA AAATTATGGA

2251  AAAATATTCT GTAACCTTTA TAAGTAGGCA TAACAGTTAT AATCATAACA

2301  TACTGTTTTT TCTTACTCCA CACAGGCATA GAGTGTCTGC TATTAATAAC

2351  TATGCTCAAA AATTGTGTAC CTTTAGCTTT TTAATTTGTA AAGGGGTTAA

2401  TAAGGAATAT TTGATGTATA GTGCCTTGAC TAGAGATCAT AATCAGCCAT

2451  ACCACATTTG TAGAGGTTTT ACTTGCTTTA AAAAACCTCC CACACCTCCC

2501  CCTGAACCTG AAACATAAAA TGAATGCAAT TGTTGTTGTT AACTTGTTTA

2551  TTGCAGCTTA TAATGGTTAC AAATAAAGCA ATAGCATCAC AAATTTCACA

2601  AATAAAGCAT TTTTTTCACT GCATTCTAGT TGTGGTTTGT CCAAACTCAT

2651  CAATGTATCT TATCATGTCT GGATCCTCTA CGCCGGACGC ATCGTGGCCG

2701  GCATCACCGG CGCCACAGGT GCGGTTGCTG GCGCCTATAT CGCCGACATC

2751  ACCGATGGGG AAGATCGGGC TCGCCACTTC GGGCTCATGA GCGCTTGTTT

2801  CGGCGTGGGT ATGGTGGCAG GCCCGTGGCC GGGGGACTGT TGGGCGCCAT

2851  CTCCTTGCAT GCACCATTCC TTGCGGCGGC GGTGCTCAAC GGCCTCAACC

2901  TACTACTGGG CTGCTTCCTA ATGCAGGAGT CGCATAAGGG AGAGCGTCGA

2951  CCGATGCCCT TGAGAGCCTT CAACCCAGTC AGCTCCTTCC GGTGGGCGCG

3001  GGGCATGACT ATCGTCGCCG CACTTATGAC TGTCTTCTTT ATCATGCAAC
```

FIG. 21 (cont'd)

```
3051 TCGTAGGACA GGTGCCGGCA GCGCTCTGGG TCATTTTCGG CGAGGACCGC
3101 TTTCGCTGGA GCGCGACGAT GATCGGCCTG TCGCTTGCGG TATTCGGAAT
3151 CTTGCACGCC CTCGCTCAAG CCTTCGTCAC TGGTCCCGCC ACCAAACGTT
3201 TCGGCGAGAA GCAGGCCATT ATCGCCGGCA TGGCGGCCGA CGCGCTGGGC
3251 TACGTCTTGC TGGCGTTCGC GACGCGAGGC TGGATGGCCT TCCCCATTAT
3301 GATTCTTCTC GCTTCCGGCG GCATCGGGAT GCCCGCGTTG CAGGCCATGC
3351 TGTCCAGGCA GGTAGATGAC GACCATCAGG GACAGCTTCA AGGATCGCTC
3401 GCGGCTCTTA CCAGCCTAAC TTCGATCACT GGACCGCTGA TCGTCACGGC
3451 GATTTATGCC GCCTCGGCGA GCACATGGAA CGGGTTGGCA TGGATTGTAG
3501 GCGCCGCCCT ATACCTTGTC TGCCTCCCCG CGTTGCGTCG CGGTGCATGG
3551 AGCCGGGCCA CCTCGACCTG AATGGAAGCC GGCGGCACCT CGCTAACGGA
3601 TTCACCACTC CAAGAATTGG AGCCAATCAA TTCTTGCGGA GAACTGTGAA
3651 TGCGCAAACC AACCCTTGGC AGAACATATC CATCGCGTCC GCCATCTCCA
3701 GCAGCCGCAC GCGGCGCATC TCGGGCCGCG TTGCTGGCGT TTTTCCATAG
3751 GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA AGTCAGAGGT
3801 GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC CCCTGGAAGC
3851 TCCCTCGTGC GCTCTCCTGT TCCGACCCTG CCGCTTACCG GATACCTGTC
3901 CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCAATGC TCACGCTGTA
3951 GGTATCTCAG TTCGGTGTAG GTCGTTCGCT CCAAGCTGGG CTGTGTGCAC
4001 GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA ACTATCGTCT
4051 TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG
4101 GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG
4151 AAGTGGTGGC CTAACTACGG CTACACTAGA AGGACAGTAT TTGGTATCTG
4201 CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT
4251 CCGGCAAACA AACCACCGCT GGTAGCGGTG GTTTTTTTGT TTGCAAGCAG
4301 CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCTT TGATCTTTTC
4351 TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA GGGATTTTGG
4401 TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCTTTT AAATTAAAAA
4451 TGAAGTTTTA AATCAATCTA AAGTATATAT GAGTAAACTT GGTCTGACAG
4501 TTACCAATGC TTAATCAGTG AGGCACCTAT CTCAGCGATC TGTCTATTTC
4551 GTTCATCCAT AGTTGCCTGA CTCCCCGTCG TGTAGATAAC TACGATACGG
4601 GAGGGCTTAC CATCTGGCCC CAGTGCTGCA ATGATACCGC GAGACCCACG
4651 CTCACCGGCT CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG
```

FIG.21 (cont'd)

```
4701  AGCGCAGAAG  TGGTCCTGCA  ACTTTATCCG  CCTCCATCCA  GTCTATTAAT
4751  TGTTGCCGGG  AAGCTAGAGT  AAGTAGTTCG  CCAGTTAATA  GTTTGCGCAA
4801  CGTTGTTGCC  ATTGCTGCAG  GCATCGTGGT  GTCACGCTCG  TCGTTTGGTA
4851  TGGCTTCATT  CAGCTCCGGT  TCCCAACGAT  CAAGGCGAGT  TACATGATCC
4901  CCCATGTTGT  GCAAAAAAGC  GGTTAGCTCC  TTCGGTCCTC  CGATCGTTGT
4951  CAGAAGTAAG  TTGGCCGCAG  TGTTATCACT  CATGGTTATG  GCAGCACTGC
5001  ATAATTCTCT  TACTGTCATG  CCATCCGTAA  GATGCTTTTC  TGTGACTGGT
5051  GAGTACTCAA  CCAAGTCATT  CTGAGAATAG  TGTATGCGGC  GACCGAGTTG
5101  CTCTTGCCCG  GCGTCAACAC  GGGATAATAC  CGCGCCACAT  AGCAGAACTT
5151  TAAAAGTGCT  CATCATTGGA  AAACGTTCTT  CGGGGCGAAA  ACTCTCAAGG
5201  ATCTTACCGC  TGTTGAGATC  CAGTTCGATG  TAACCCACTC  GTGCACCCAA
5251  CTGATCTTCA  GCATCTTTTA  CTTTCACCAG  CGTTTCTGGG  TGAGCAAAAA
5301  CAGGAAGGCA  AAATGCCGCA  AAAAAGGGAA  TAAGGGCGAC  ACGGAAATGT
5351  TGAATACTCA  TACTCTTCCT  TTTTCAATAT  TATTGAAGCA  TTTATCAGGG
5401  TTATTGTCTC  ATGAGCGGAT  ACATATTTGA  ATGTATTTAG  AAAAATAAAC
5451  AAATAGGGGT  TCCGCGCACA  TTTCCCCGAA  AAGTGCCACC  TGACGTCTAA
5501  GAAACCATTA  TTATCATGAC  ATTAACCTAT  AAAAATAGGC  GTATCACGAG
5551  GCCCTTTCGT  CTTCAA
```

pBG395
:BG368 backbone
:soluble T4#10
:AA #3 = LYS
:first: 131 AA of T4

FIG. 22

pg395.seq Length 5413

```
  1  GAATTAATTC CAGCTTGCTG TGGAATGTGT GTCAGTTAGG GTGTGGAAAG

 51  TCCCCAGGCT CCCCAGCAGG CAGAAGTATG CAAAGCATGC ATCTCAATTA

101  GTCAGCAACC AGGTGTGGAA AGTCCCCAGG CTCCCCAGCA GGCAGAAGTA

151  TGCAAAGCAT GCATCTCAAT TAGTCAGCAA CCATAGTCCC GCCCCTAACT

201  CCGCCCATCC CGCCCCTAAC TCCGCCCAGT TCCGCCCATT CTCCGCCCCA

251  TGGCTGACTA ATTTTTTTTA TTTATGCAGA GGCCGAGGCC GCCTCGGCCT

301  CTGAGCTATT CCAGAAGTAG TGAGGAGGCT TTTTTGGAGG GGTCCTCCTC

351  GTATAGAAAC TCGGACCACT CTGAGACGAA GGCTCGCGTC CAGGCCAGCA

401  CGAAGGAGGC TAAGTGGGAG GGGTAGCGGT CGTTGTCCAC TAGGGGGTCC

451  ACTCGCTCCA GGGTGTGAAG ACACATGTCG CCCTCTTCGG CATCAAGGAA

501  GGTGATTGGT TTATAGGTGT AGGCCACGTG ACCGGGTGTT CCTGAAGGGG

551  GGCTATAAAA GGGGGTGGGG GCGCGTTCGT CCTCACTCTC TTCCGCATCG

601  CTGTCTGCGA GGGCCAGCTG TTGGGCTCGC GGTTGAGGAC AAACTCTTCG

651  CGGTCTTTCC AGTACTCTTG GATCGGAAAC CCGTCGGCCT CCGAACGGTA

701  CTCCGCCACC GAGGGACCTG AGCGAGTCCG CATCGACCGG ATCGGAAAAC

751  CTCTCGAGAA AGGCGTCTAA CCAGTCACAG TCGCAAGGTA GGCTGAGCAC

801  CGTGGCGGGC GGCAGCGGGT GGCGGTCGGG GTTGTTTCTG GCGGAGGTGC

851  TGCTGATGAT GTAATTAAAG TAGGCGGTCT TGAGACGGCG GATGGTCGAG

901  GTGAGGTGTG GCAGGCTTGA GATCGATCTG GCCATACACT TGAGTGACAA

951  TGACATCCAC TTTGCCTTTC TCTCCACAGG TGTCCACTCC CAGGTCCAAC

1001  TGGATCCAAG CTTCGACTCG AGGAATTCCC CGAAGGAACA AAGCACCCTC

1051  CCCACTGGGC TCCTGGTTGC AGAGCTCCAA GTCCTCACAC AGATACGCCT

1101  GTTTGAGAAG CAGCGGGCAA GAAAGACGCA AGCCCAGAGG CCCTGCCATT

1151  TCTGTGGGCT CAGGTCCCTA CTGGCTCAGG CCCCTGCCTC CCTCGGCAAG

1201  GCCACAATGA ACCGGGGAGT CCCTTTTAGG CACTTGCTTC TGGTGCTGCA

1251  ACTGGCGCTC CTCCCAGCAG CCACTCAGGG AAAGAAAGTG GTGCTGGGCA

1301  AAAAAGGGGA TACAGTGGAA CTGACCTGTA CAGCTTCCCA GAAGAAGAGC

1351  ATACAATTCC ACTGGAAAAA CTCCAACCAG ATAAAGATTC TGGGAAATCA

1401  GGGCTCCTTC TTAACTAAAG GTCCATCCAA GCTGAATGAT CGCGCTGACT
```

FIG. 22(cont'd)

```
1451:  CAAGAAGAAG  CTTGTGGGAC  CAAGGAAACT  TTCCCCTGAT  CATCAAGAAT
1501:  CTTAAGATAG  AAGACTCAGA  TACTTACATC  TGTGAAGTGG  AGGACCAGAA
1551:  GGAGGAGGTG  CAATTGCTAG  TGTTCGGATT  GACTGCCAAC  TCTGACACCC
1601:  ACCTGCTTCA  GGGGCAGAGC  CTGACCCTGA  CCTTGGAGAG  CCCCCCGGGT
1651:  AGTAGCCCCT  CAGTGCAATG  AGATCTTTGT  GAAGGAACCT  TACTTCTGTG
1701:  GTGTGACATA  ATTGGACAAA  CTACCTACAG  AGATTTAAAG  CTCTAAGGTA
1751:  AATATAAAAT  TTTTAAGTGT  ATAATGTGTT  AAACTACTGA  TTCTAATTGT
1801:  TTGTGTATTT  TAGATTCCAA  CCTATGGAAC  TGATGAATGG  GAGCAGTGGT
1851:  GGAATGCCTT  TAATGAGGAA  AACCTGTTTT  GCTCAGAAGA  AATGCCATCT
1901:  AGTGATGATG  AGGCTACTGC  TGACTCTCAA  CATTCTACTC  CTCCAAAAAA
1951:  GAAGAGAAAG  GTAGAAGACC  CCAAGGACTT  TCCTTCAGAA  TTGCTAAGTT
2001:  TTTTGAGTCA  TGCTGTGTTT  AGTAATAGAA  CTCTTGCTTG  CTTTGCTATT
2051:  TACACCACAA  AGGAAAAAGC  TGCACTGCTA  TACAAGAAAA  TTATGGAAAA
2101:  ATATTCTGTA  ACCTTTATAA  GTAGGCATAA  CAGTTATAAT  CATAACATAC
2151:  TGTTTTTTCT  TACTCCACAC  AGGCATAGAG  TGTCTGCTAT  TAATAACTAT
2201:  GCTCAAAAAT  TGTGTACCTT  TAGCTTTTTA  ATTTGTAAAG  GGGTTAATAA
2251:  GGAATATTTG  ATGTATAGTG  CCTTGACTAG  AGATCATAAT  CAGCCATACC
2301:  ACATTTGTAG  AGGTTTTACT  TGCTTTAAAA  AACCTCCCAC  ACCTCCCCCT
2351:  GAACCTGAAA  CATAAAATGA  ATGCAATTGT  TGTTGTTAAC  TTGTTTATTG
2401:  CAGCTTATAA  TGGTTACAAA  TAAAGCAATA  GCATCACAAA  TTTCACAAAT
2451:  AAAGCATTTT  TTTCACTGCA  TTCTAGTTGT  GGTTTGTCCA  AACTCATCAA
2501:  TGTATCTTAT  CATGTCTGGA  TCCTCTACGC  CGGACGCATC  GTGGCCGGCA
2551:  TCACCGGCGC  CACAGGTGCG  GTTGCTGGCG  CCTATATCGC  CGACATCACC
2601:  GATGGGGAAG  ATCGGGCTCG  CCACTTCGGG  CTCATGAGCG  CTTGTTTCGG
2651:  CGTGGGTATG  GTGGCAGGCC  CGTGGCCGGG  GGACTGTTGG  GCGCCATCTC
2701:  CTTGCATGCA  CCATTCCTTG  CGGCGGCGGT  GCTCAACGGC  CTCAACCTAC
2751:  TACTGGGCTG  CTTCCTAATG  CAGGAGTCGC  ATAAGGGAGA  GCGTCGACCG
2801:  ATGCCCTTGA  GAGCCTTCAA  CCCAGTCAGC  TCCTTCCGGT  GGGCGCGGGG
2851:  CATGACTATC  GTCGCCGCAC  TTATGACTGT  CTTCTTTATC  ATGCAACTCG
2901:  TAGGACAGGT  GCCGGCAGCG  CTCTGGGTCA  TTTTCGGCGA  GGACCGCTTT
2951:  CGCTGGAGCG  CGACGATGAT  CGGCCTGTCG  CTTGCGGTAT  TCGGAATCTT
3001:  GCACGCCCTC  GCTCAAGCCT  TCGTCACTGG  TCCCGCCACC  AAACGTTTCG
3051:  GCGAGAAGCA  GGCCATTATC  GCCGGCATGG  CGGCCGACGC  GCTGGGCTAC
```

FIG.22(cont'd)

```
3101  GTCTTGCTGG CGTTCGCGAC GCGAGGCTGG ATGGCCTTCC CCATTATGAT
3151  TCTTCTCGCT TCCGGCGGCA TCGGGATGCC CGCGTTGCAG GCCATGCTGT
3201  CCAGGCAGGT AGATGACGAC CATCAGGGAC AGCTTCAAGG ATCGCTCGCG
3251  GCTCTTACCA GCCTAACTTC GATCACTGGA CCGCTGATCG TCACGGCGAT
3301  TTATGCCGCC TCGGCGAGCA CATGGAACGG GTTGGCATGG ATTGTAGGCG
3351  CCGCCCTATA CCTTGTCTGC CTCCCCGCGT TGCGTCGCGG TGCATGGAGC
3401  CGGGCCACCT CGACCTGAAT GGAAGCCGGC GGCACCTCGC TAACGGATTC
3451  ACCACTCCAA GAATTGGAGC CAATCAATTC TTGCGGAGAA CTGTGAATGC
3501  GCAAACCAAC CCTTGGCAGA ACATATCCAT CGCGTCCGCC ATCTCCAGCA
3551  GCCGCACGCG GCGCATCTCG GGCCGCGTTG CTGGCGTTTT TCCATAGGCT
3601  CCGCCCCCCT GACGAGCATC ACAAAAATCG ACGCTCAAGT CAGAGGTGGC
3651  GAAACCCGAC AGGACTATAA AGATACCAGG CGTTTCCCCC TGGAAGCTCC
3701  CTCGTGCGCT CTCCTGTTCC GACCCTGCCG CTTACCGGAT ACCTGTCCGC
3751  CTTTCTCCCT TCGGGAAGCG TGGCGCTTTC TCAATGCTCA CGCTGTAGGT
3801  ATCTCAGTTC GGTGTAGGTC GTTCGCTCCA AGCTGGGCTG TGTGCACGAA
3851  CCCCCCGTTC AGCCCGACCG CTGCGCCTTA TCCGGTAACT ATCGTCTTGA
3901  GTCCAACCCG GTAAGACACG ACTTATCGCC ACTGGCAGCA GCCACTGGTA
3951  ACAGGATTAG CAGAGCGAGG TATGTAGGCG GTGCTACAGA GTTCTTGAAG
4001  TGGTGGCCTA ACTACGGCTA CACTAGAAGG ACAGTATTTG GTATCTGCGC
4051  TCTGCTGAAG CCAGTTACCT TCGGAAAAAG AGTTGGTAGC TCTTGATCCG
4101  GCAAACAAAC CACCGCTGGT AGCGGTGGTT TTTTGTTTG CAAGCAGCAG
4151  ATTACGCGCA GAAAAAAGG ATCTCAAGAA GATCCTTTGA TCTTTTCTAC
4201  GGGGTCTGAC GCTCAGTGGA ACGAAAACTC ACGTTAAGGG ATTTTGGTCA
4251  TGAGATTATC AAAAAGGATC TTCACCTAGA TCCTTTTAAA TTAAAAATGA
4301  AGTTTTAAAT CAATCTAAAG TATATATGAG TAAACTTGGT CTGACAGTTA
4351  CCAATGCTTA ATCAGTGAGG CACCTATCTC AGCGATCTGT CTATTTCGTT
4401  CATCCATAGT TGCCTGACTC CCCGTCGTGT AGATAACTAC GATACGGGAG
4451  GGCTTACCAT CTGGCCCCAG TGCTGCAATG ATACCGCGAG ACCCACGCTC
4501  ACCGGCTCCA GATTTATCAG CAATAAACCA GCCAGCCGGA AGGGCCGAGC
4551  GCAGAAGTGG TCCTGCAACT TTATCCGCCT CCATCCAGTC TATTAATTGT
4601  TGCCGGGAAG CTAGAGTAAG TAGTTCGCCA GTTAATAGTT TGCGCAACGT
4651  TGTTGCCATT GCTGCAGGCA TCGTGGTGTC ACGCTCGTCG TTTGGTATGG
4701  CTTCATTCAG CTCCGGTTCC CAACGATCAA GGCGAGTTAC ATGATCCCCC
```

FIG. 22(cont'd)

```
4751  ATGTTGTGCA AAAAAGCGGT TAGCTCCTTC GGTCCTCCGA TCGTTGTCAG

4801  AAGTAAGTTG GCCGCAGTGT TATCACTCAT GGTTATGGCA GCACTGCATA

4851  ATTCTCTTAC TGTCATGCCA TCCGTAAGAT GCTTTTCTGT GACTGGTGAG

4901  TACTCAACCA AGTCATTCTG AGAATAGTGT ATGCGGCGAC CGAGTTGCTC

4951  TTGCCCGGCG TCAACACGGG ATAATACCGC GCCACATAGC AGAACTTTAA

5001  AAGTGCTCAT CATTGGAAAA CGTTCTTCGG GGCGAAAACT CTCAAGGATC

5051  TTACCGCTGT TGAGATCCAG TTCGATGTAA CCCACTCGTG CACCCAACTG

5101  ATCTTCAGCA TCTTTTACTT TCACCAGCGT TTCTGGGTGA GCAAAAACAG

5151  GAAGGCAAAA TGCCGCAAAA AAGGGAATAA GGGCGACACG GAAATGTTGA

5201  ATACTCATAC TCTTCCTTTT TCAATATTAT TGAAGCATTT ATCAGGGTTA

5251  TTGTCTCATG AGCGGATACA TATTTGAATG TATTTAGAAA AATAAACAAA

5301  TAGGGGTTCC GCGCACATTT CCCCGAAAAG TGCCACCTGA CGTCTAAGAA

5351  ACCATTATTA TCATGACATT AACCTATAAA AATAGGCGTA TCACGAGGCC

5401  CTTTCGTCTT CAA
```

FIG. 23
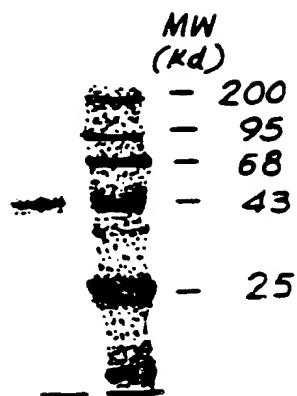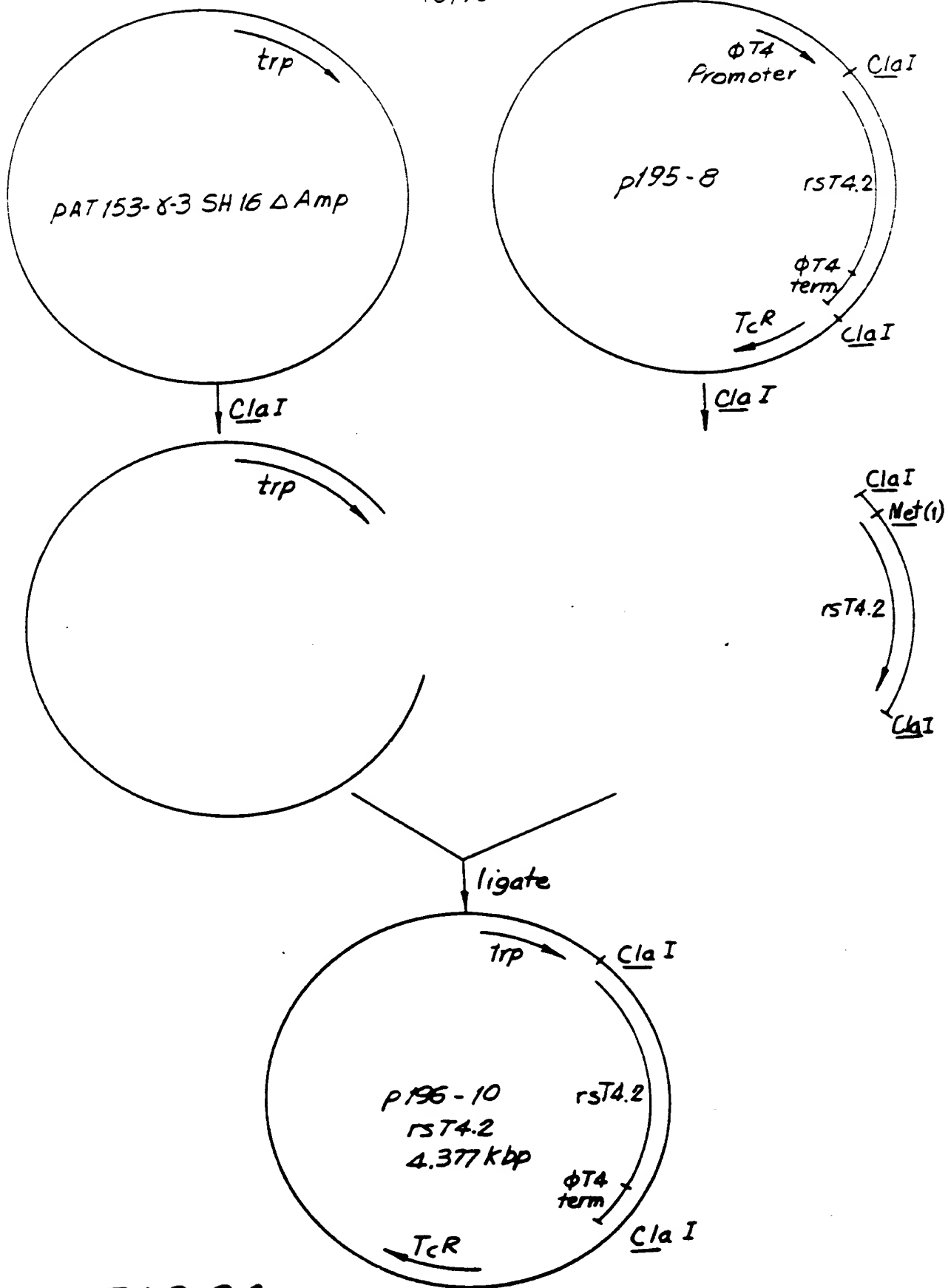
FIG.24
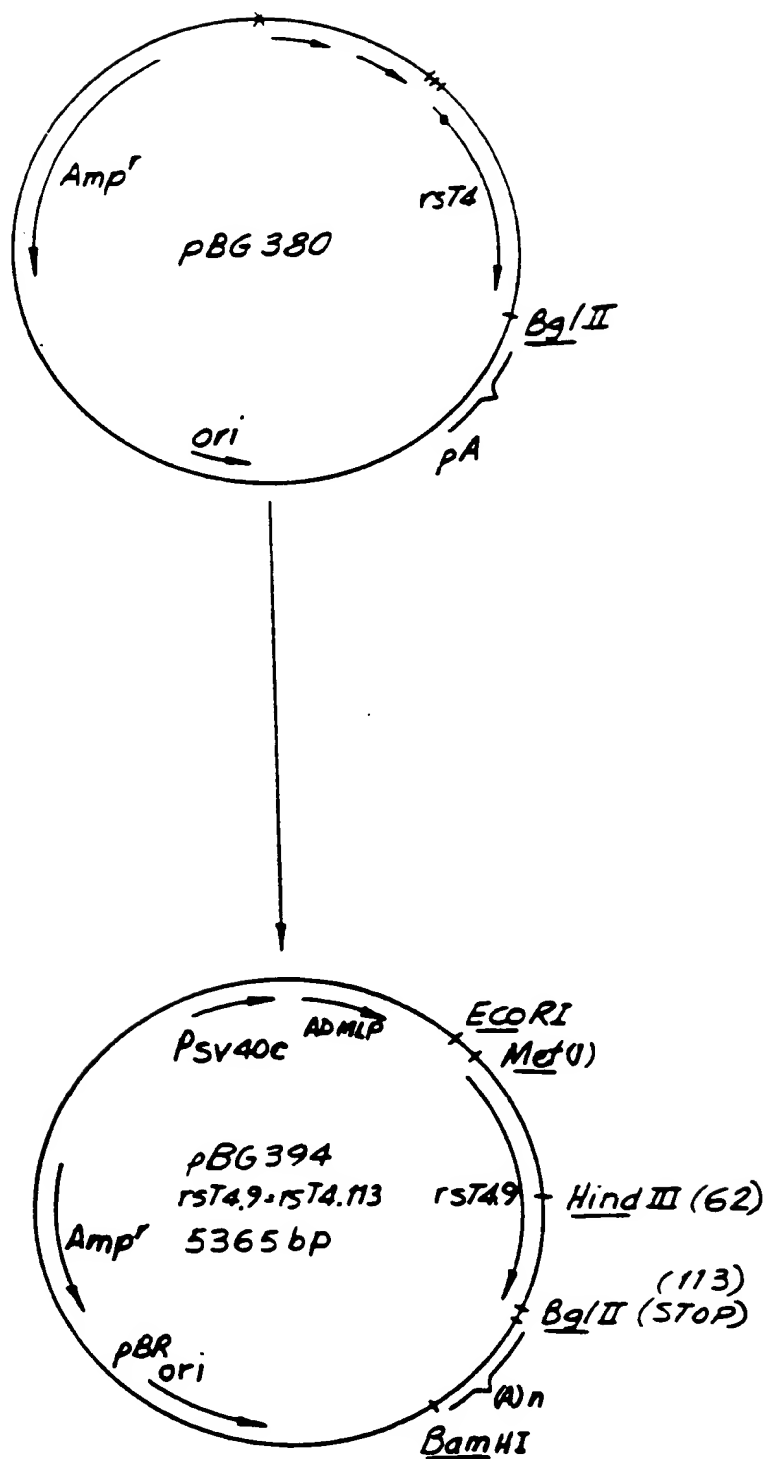
FIG.25
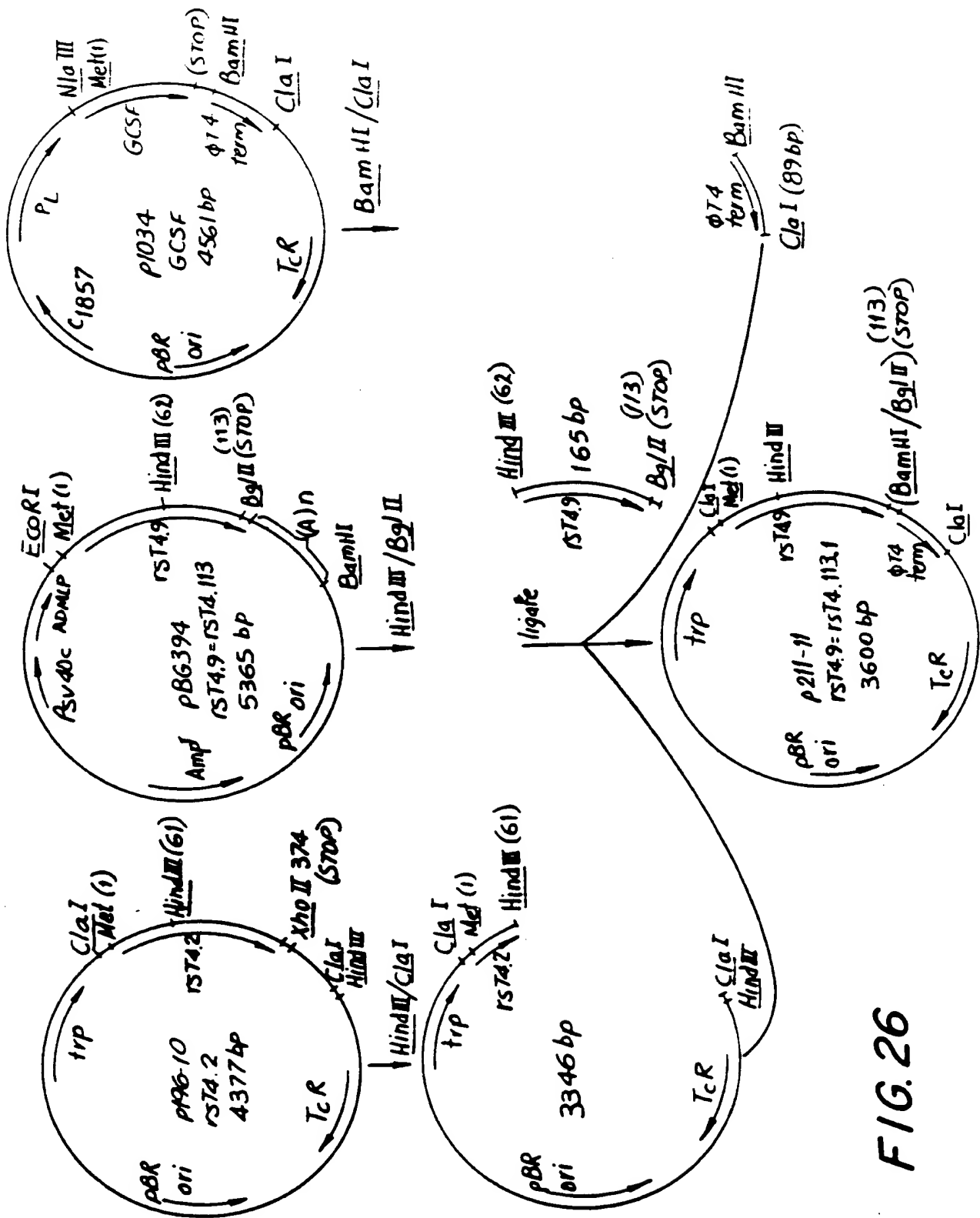
FIG. 26

FIG. 27
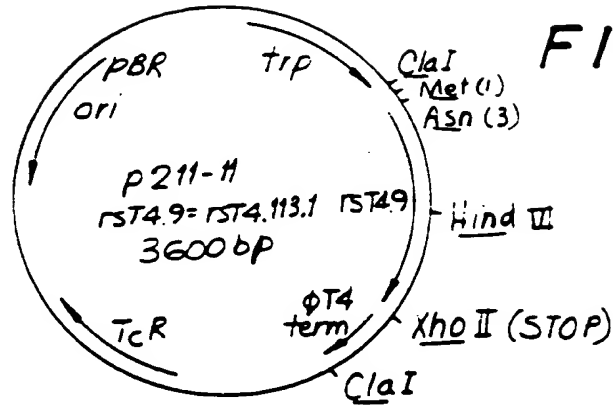
Site directed mutagenesis
to change an Asn at amino
acid position #3 to a Lys,
using T4-66
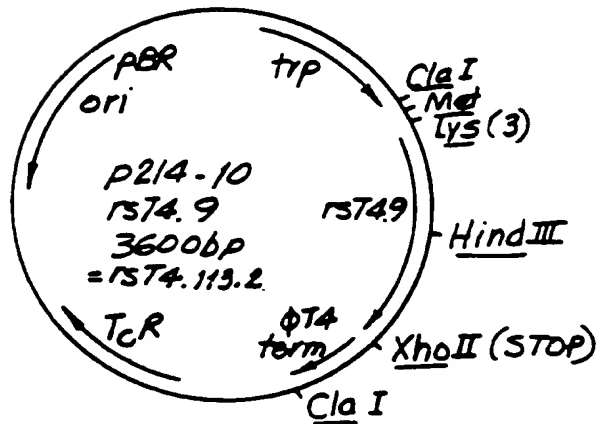
Site directed mutagenesis
to delete Gln and Gly at
amino acid positions #1, #2,
using T4AID-87
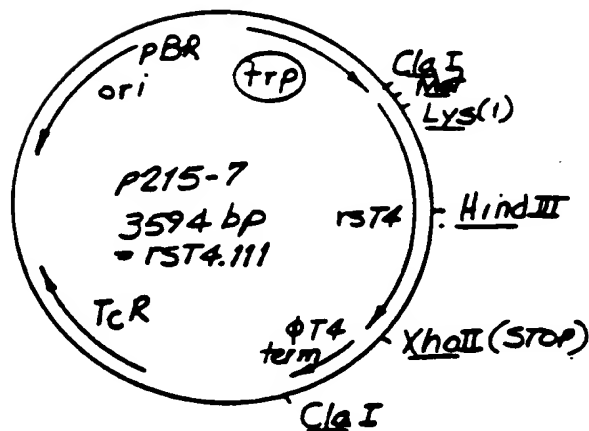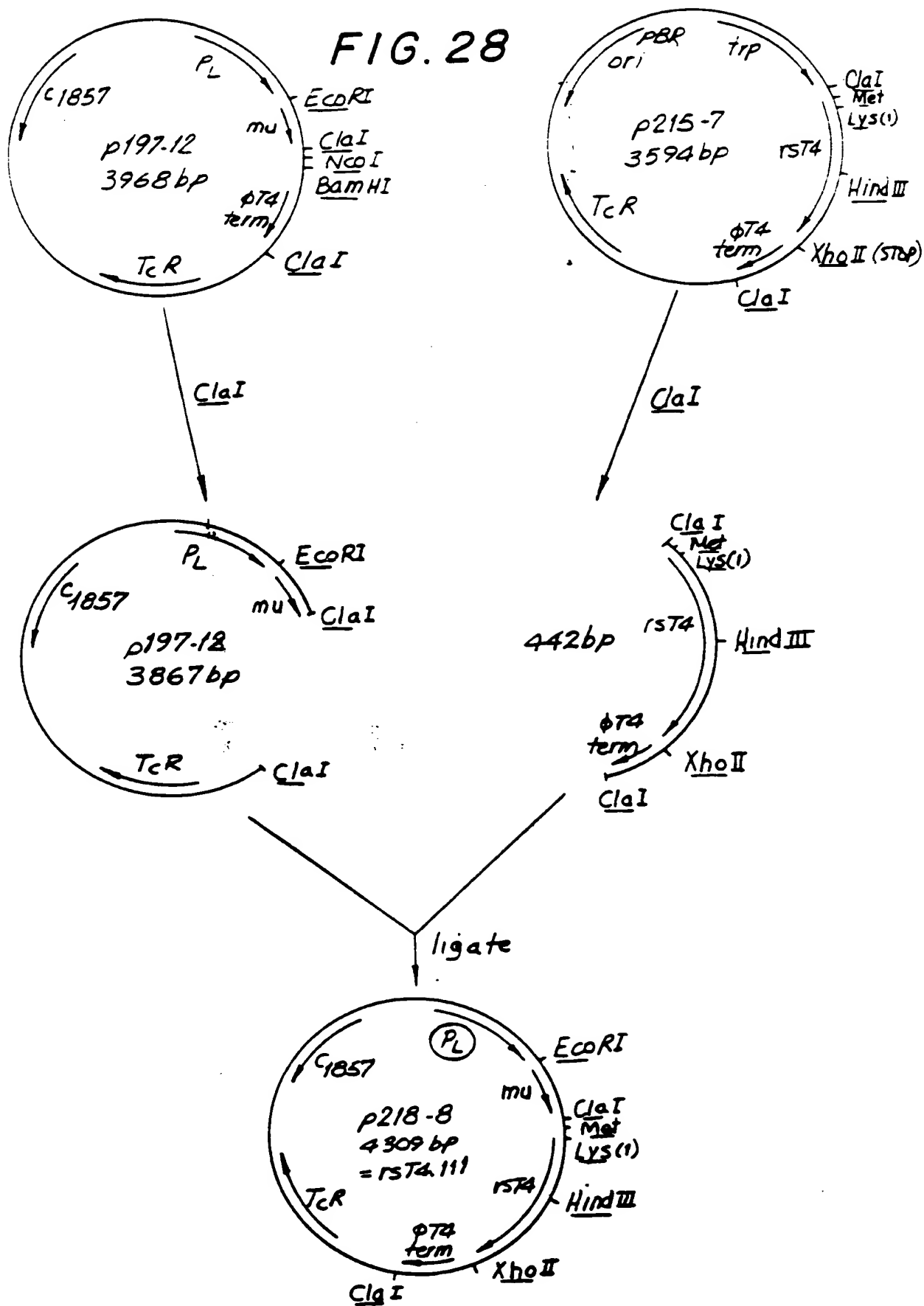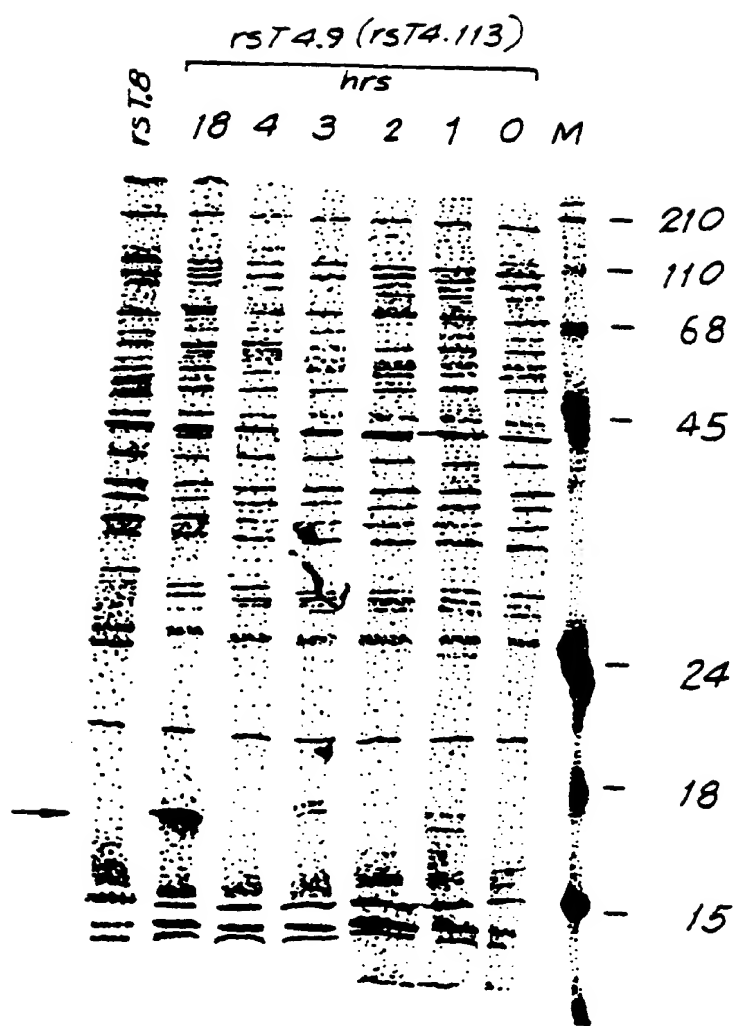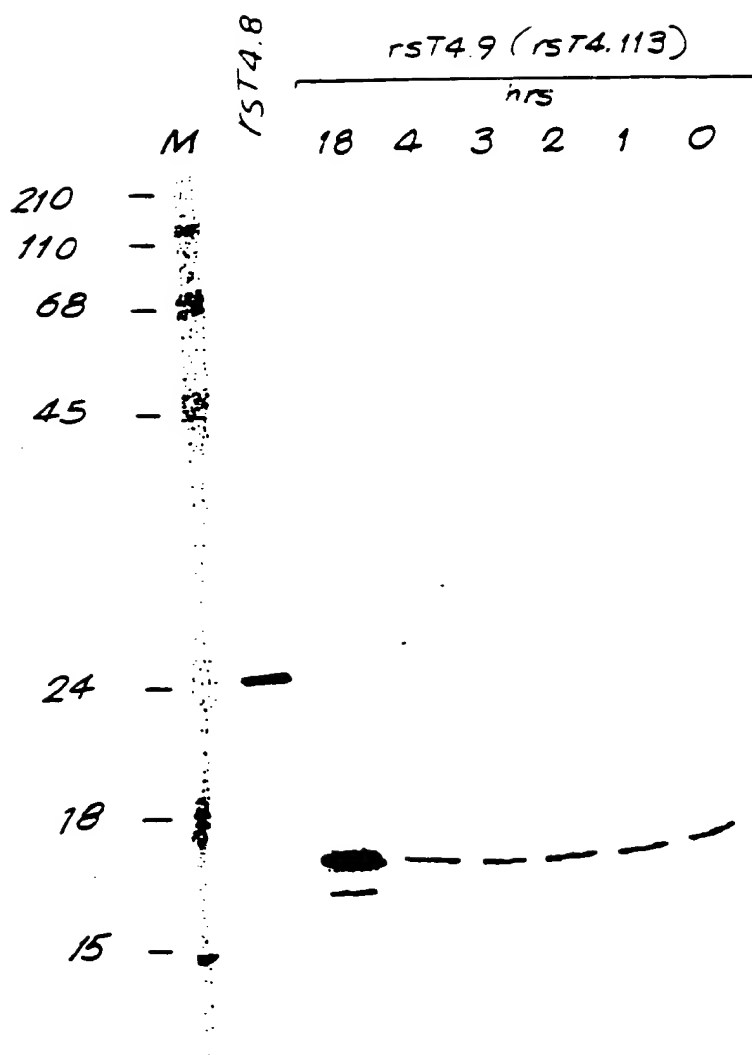
FIG. 28

FIG. 29A

FIG. 29B

F I G. 30

81/93

FIG. 31

a) b) c)

Absorbance (280nm)

Time
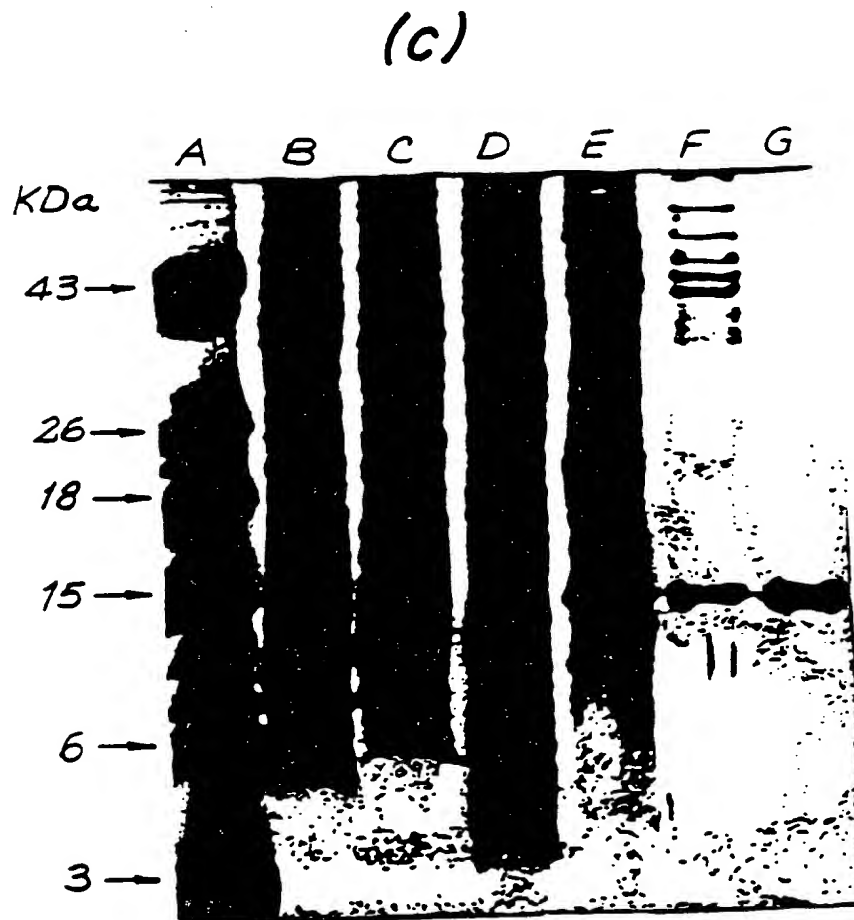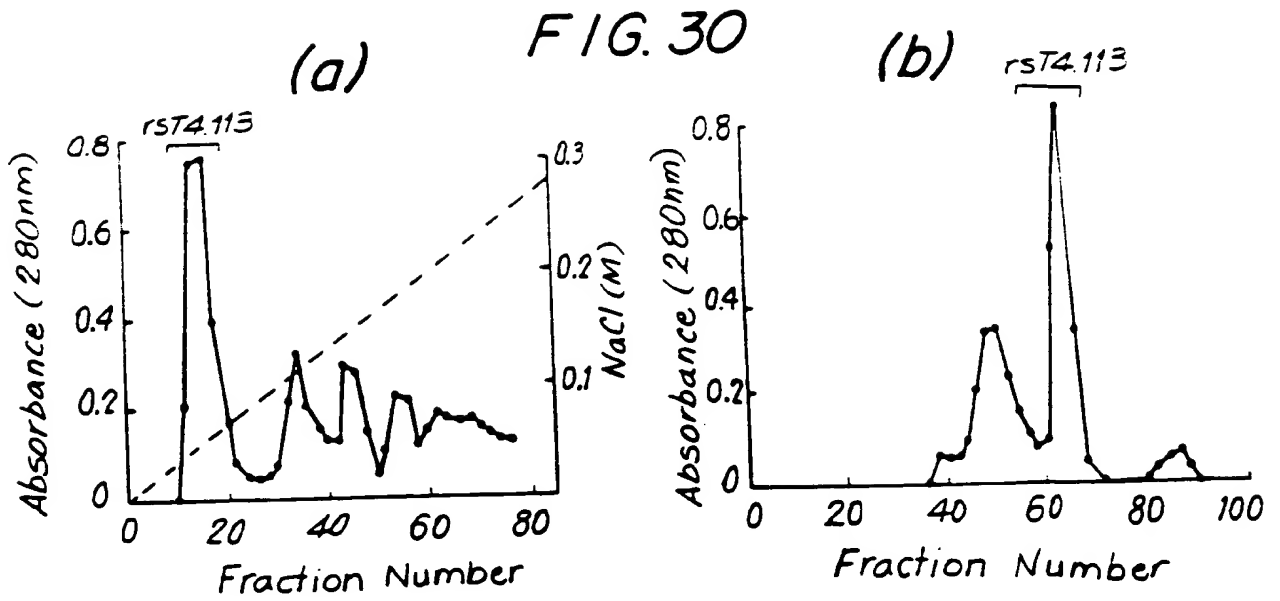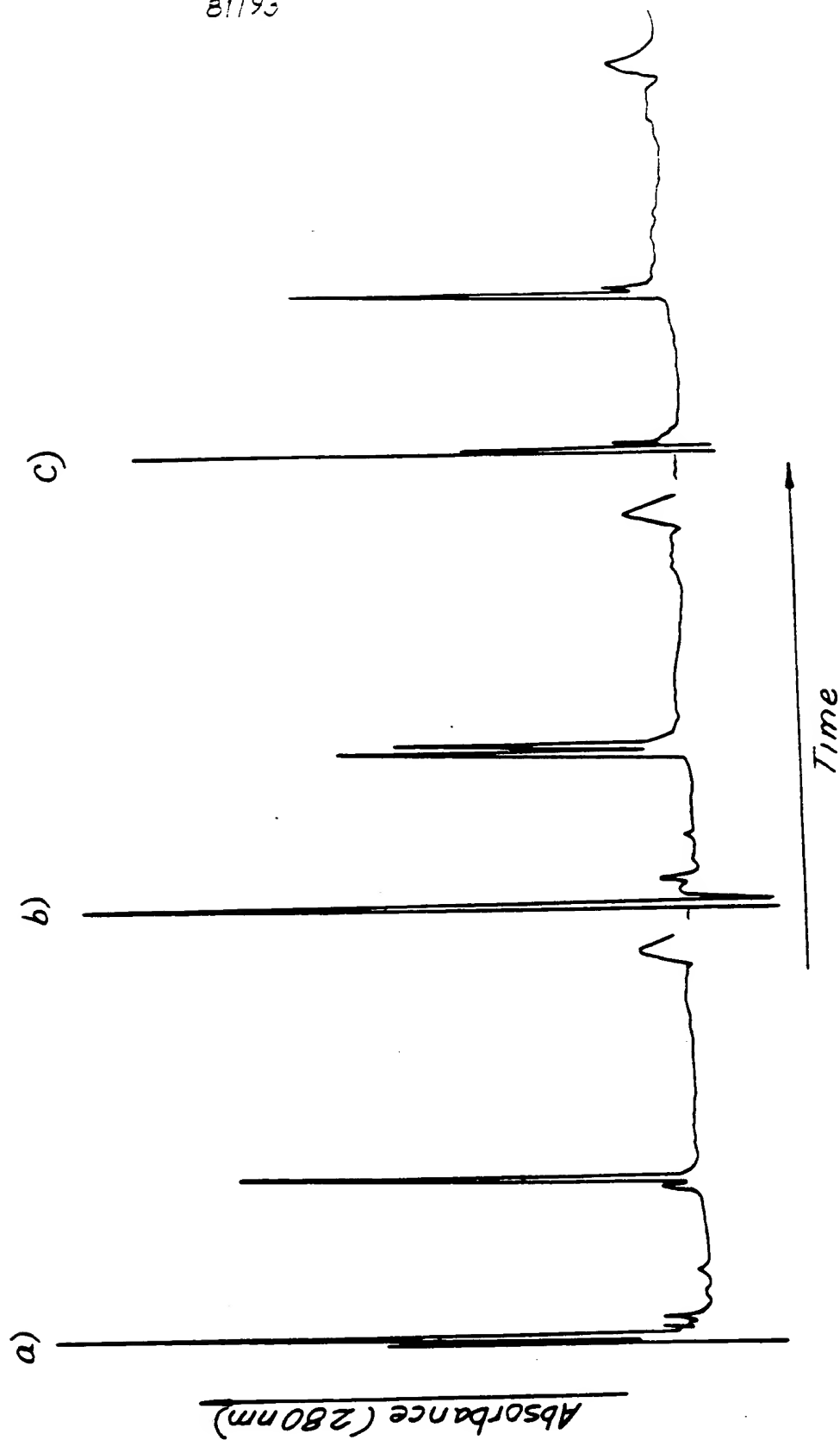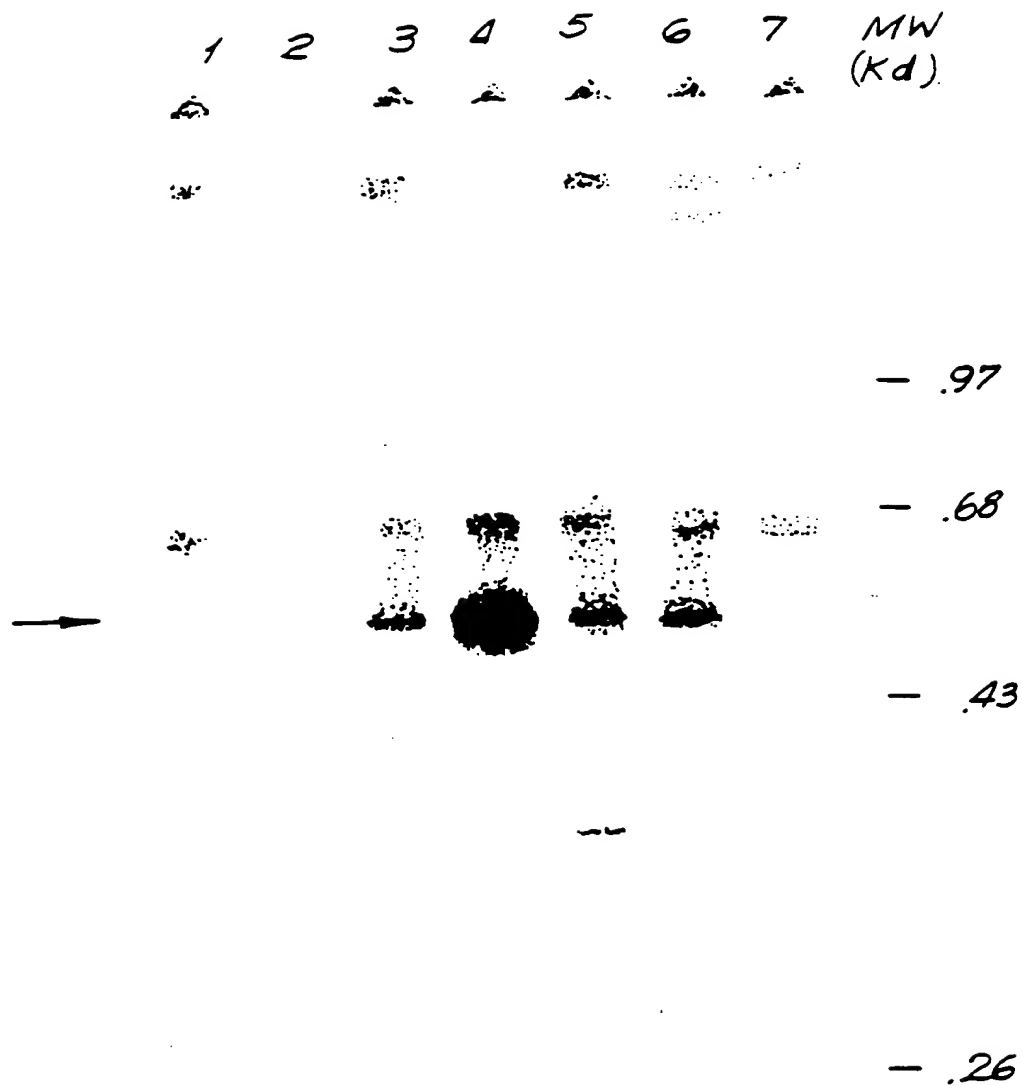
FIG. 32
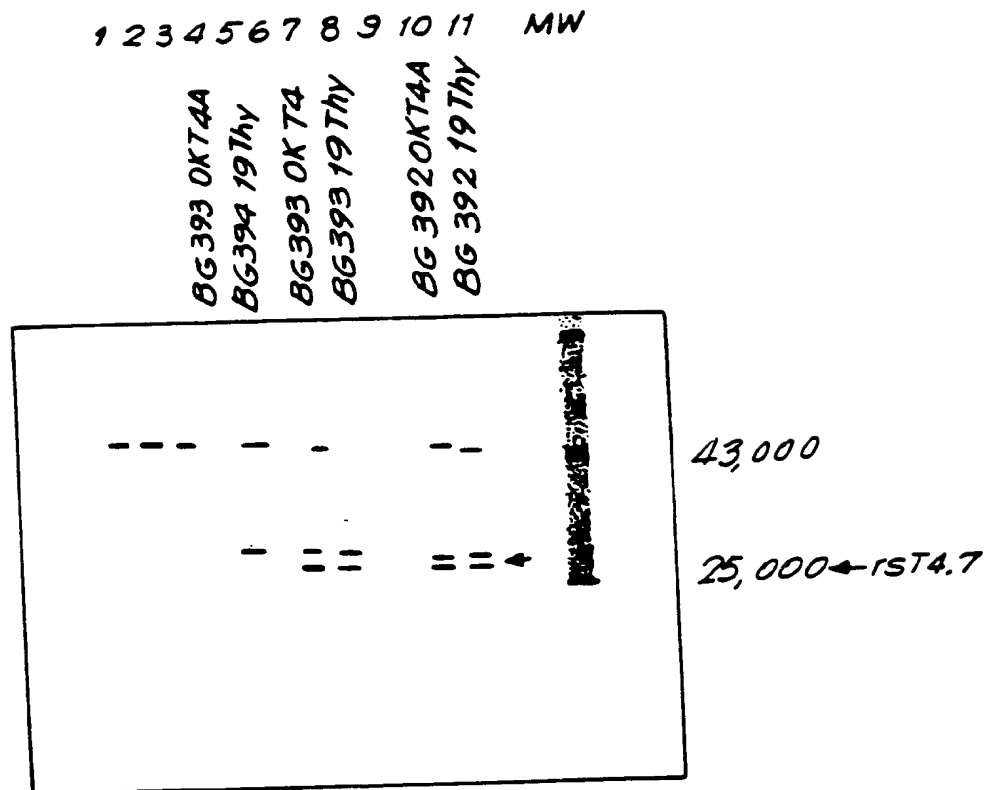
F I G. 33

FIG.34
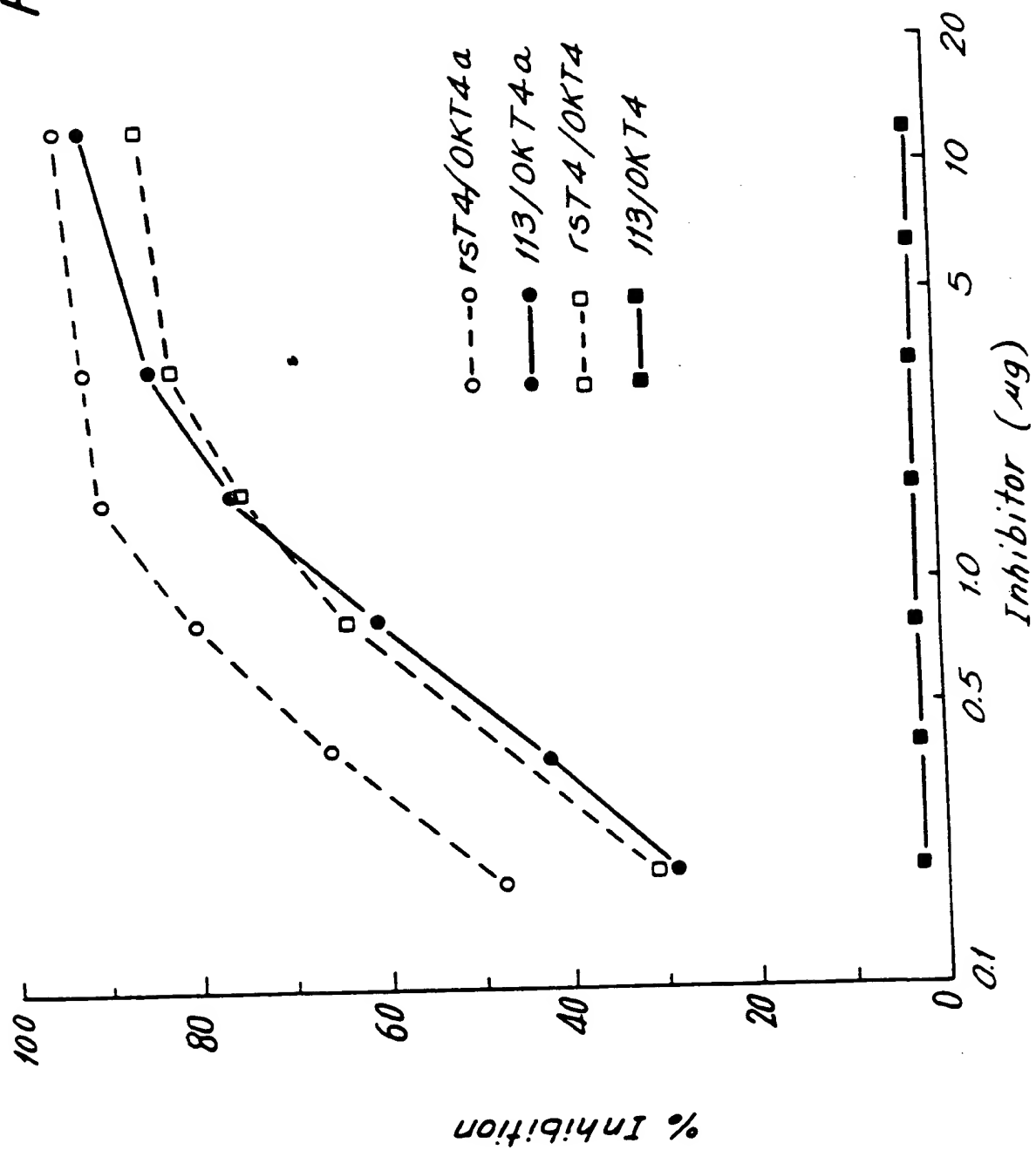
FIG. 35
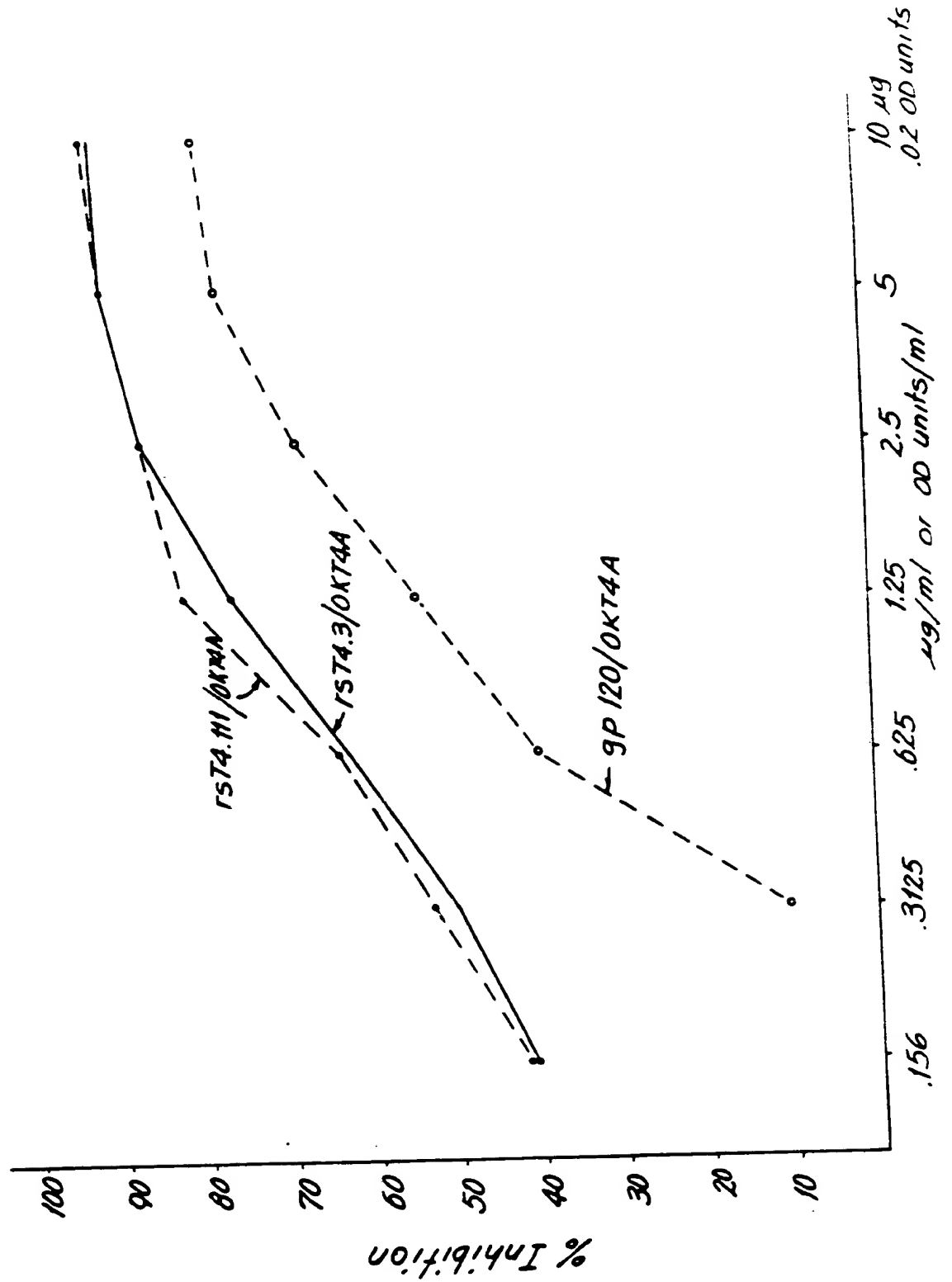
FIG.36
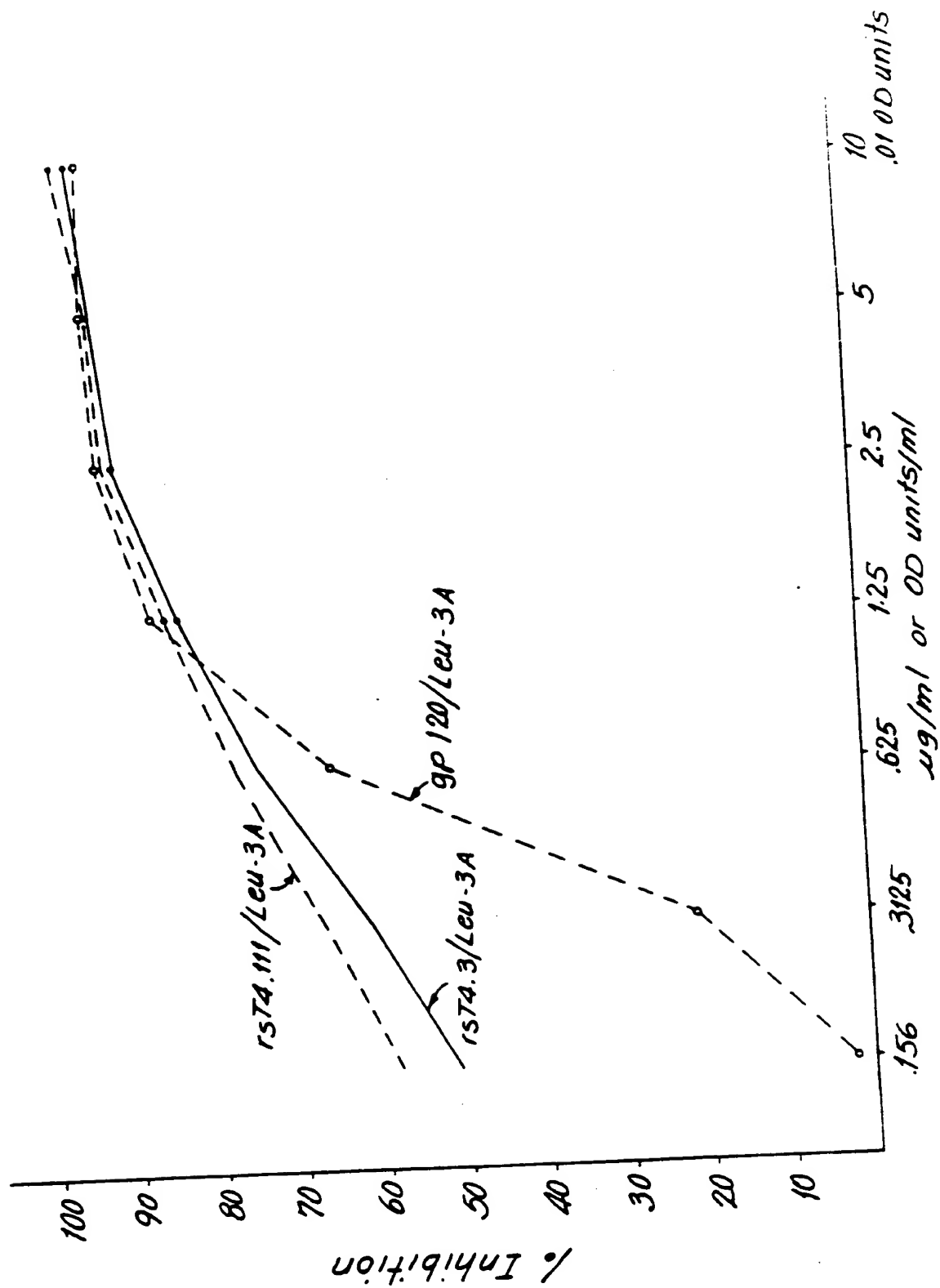
FIG. 37
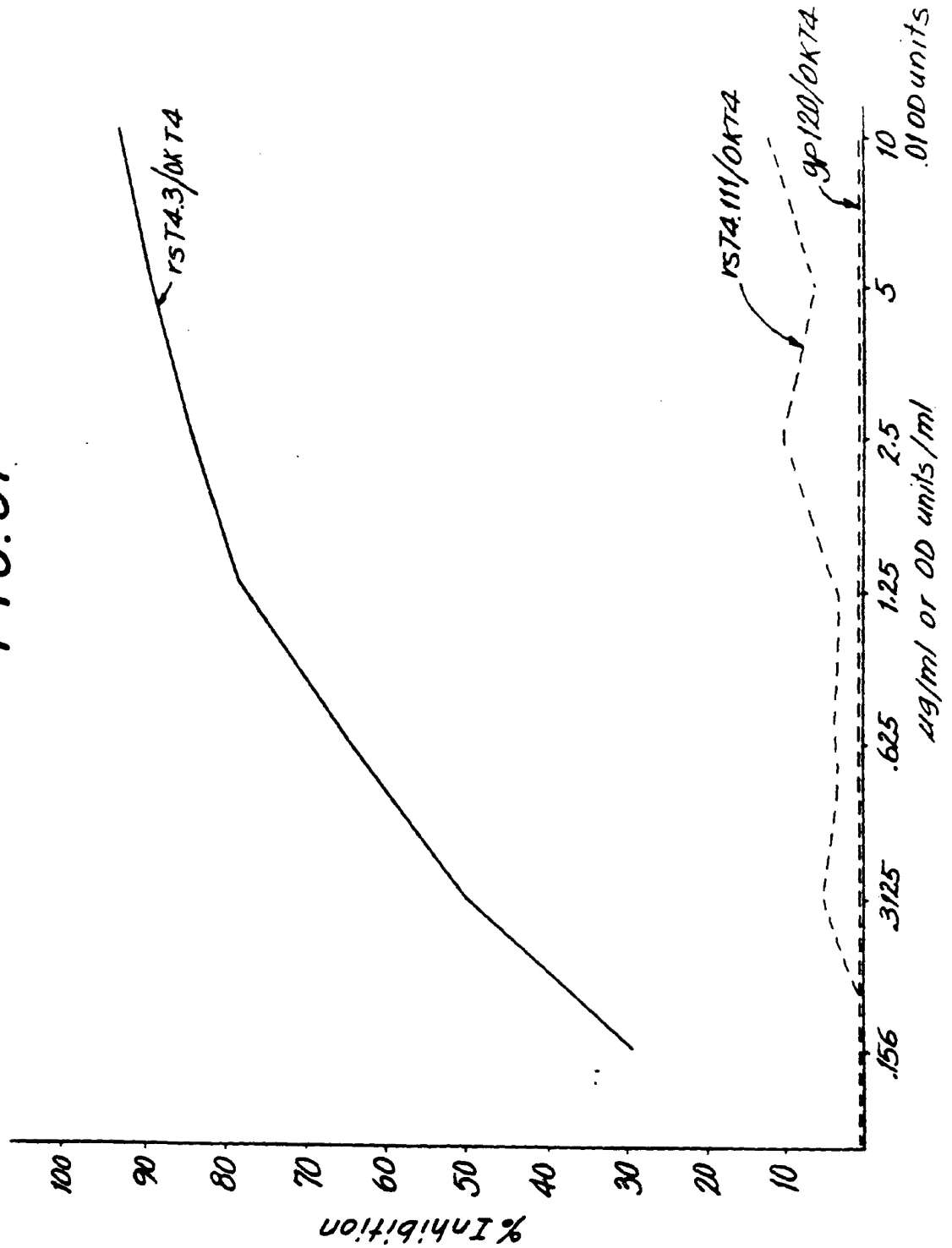
FIG. 38

Competition ELISA for rsT4.113

FIG.39

P24 ANTIGEN RIA

△ 0.05 µg/ml rST4
○ 0.5 µg/ml rST4
□ 5 µg/ml rST4
■ BUFFER

P24 ag (ng/ml)

DAYS
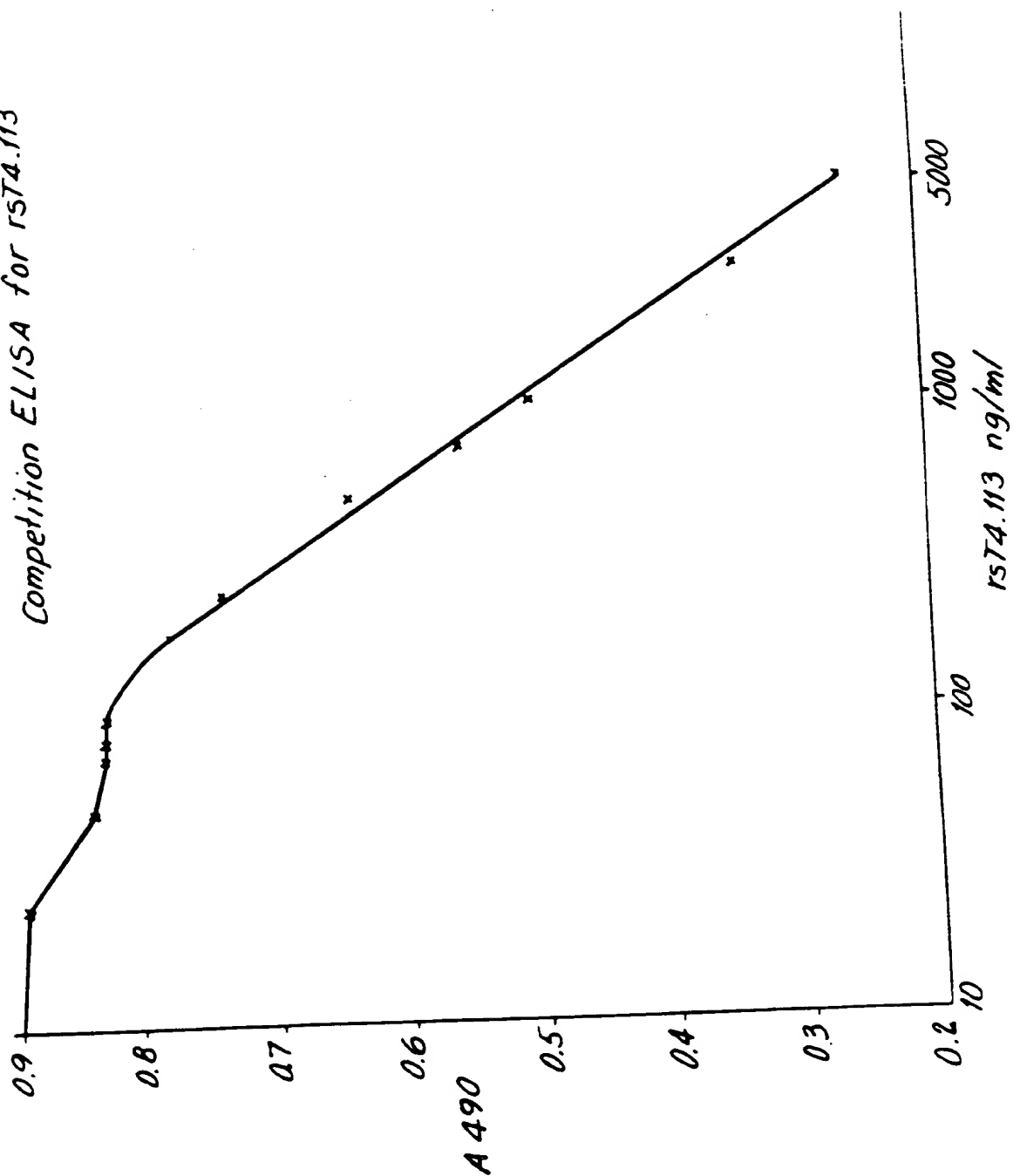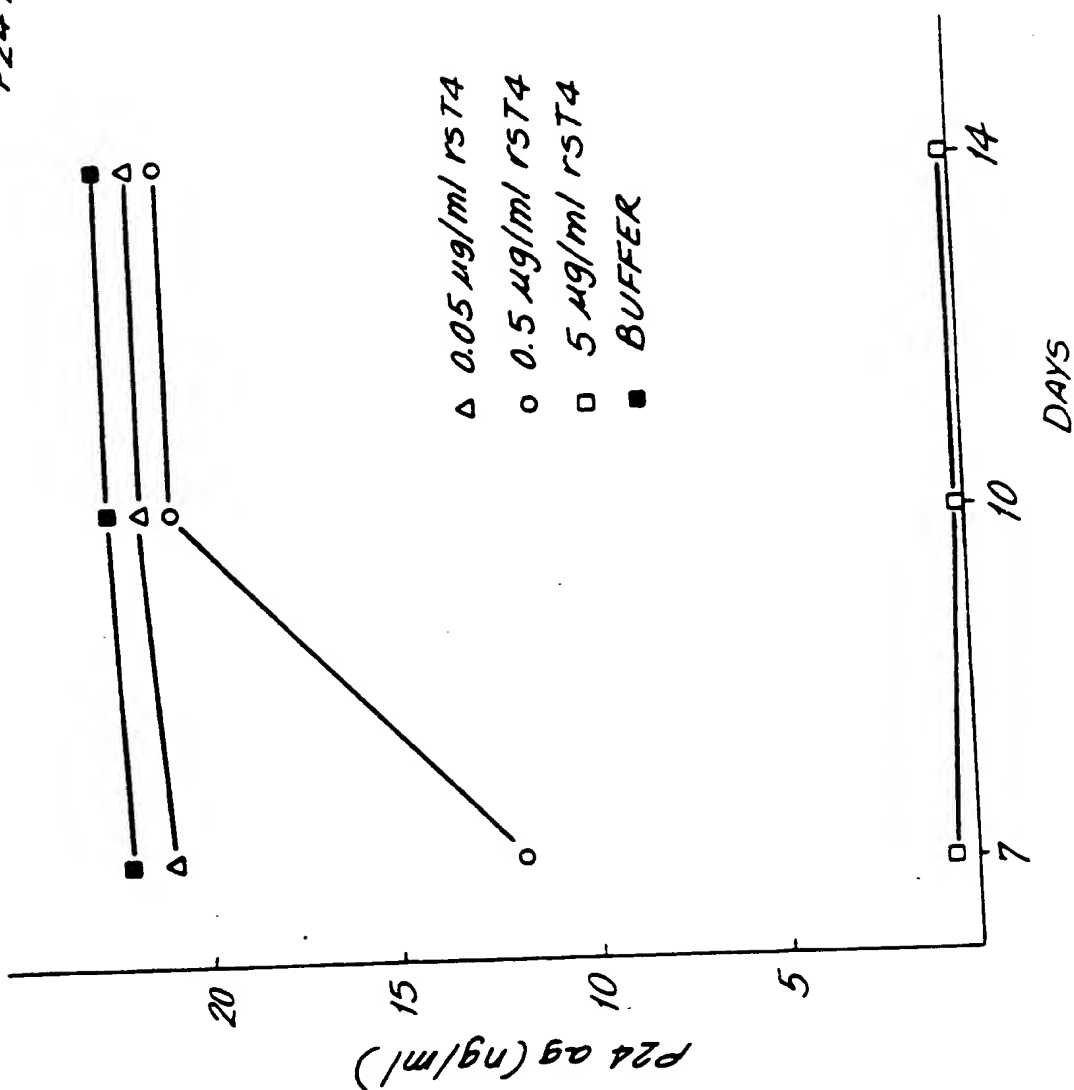
FIG. 40

C8166 CELL FUSION ASSAY
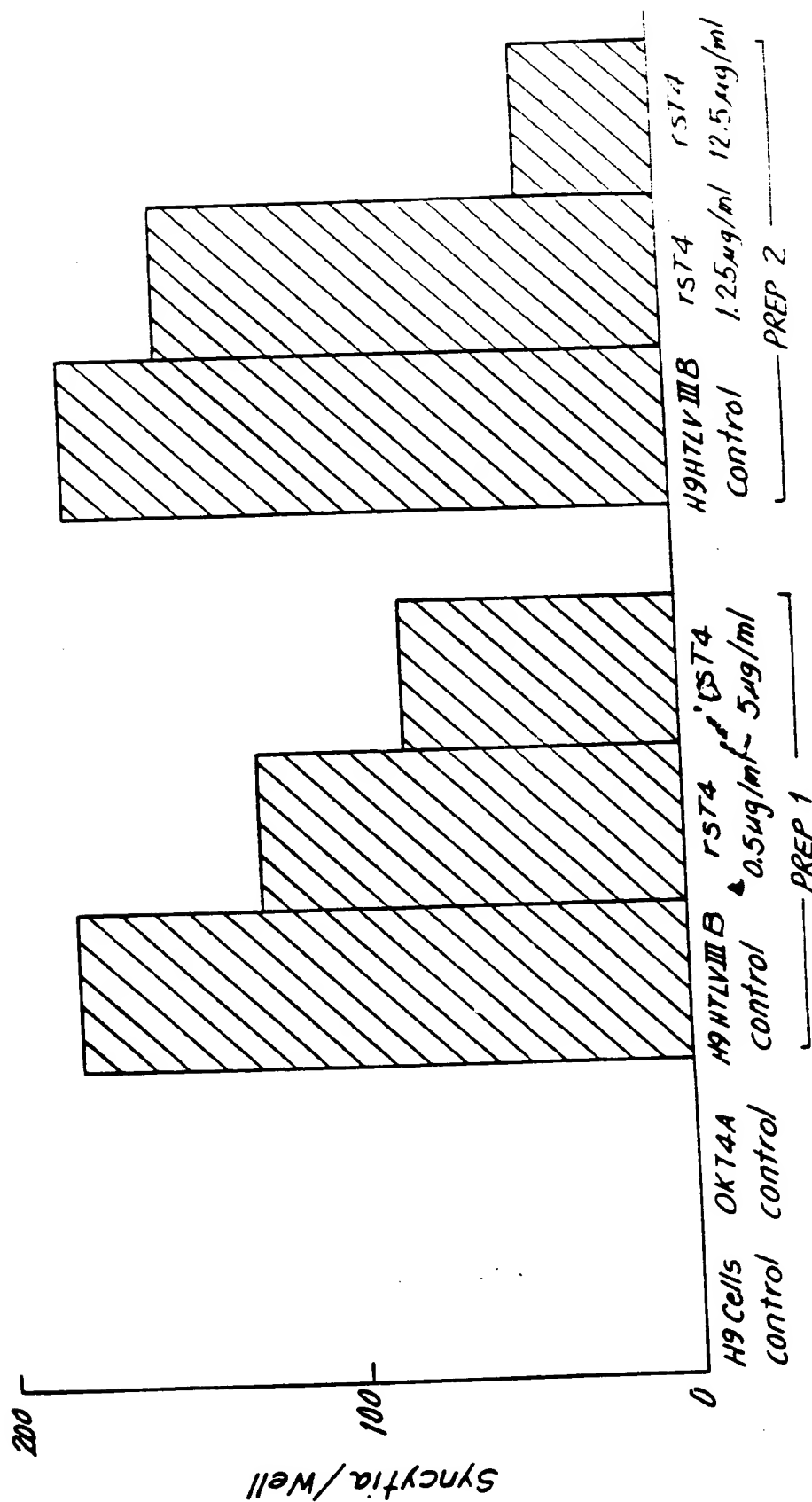
FIG. 41
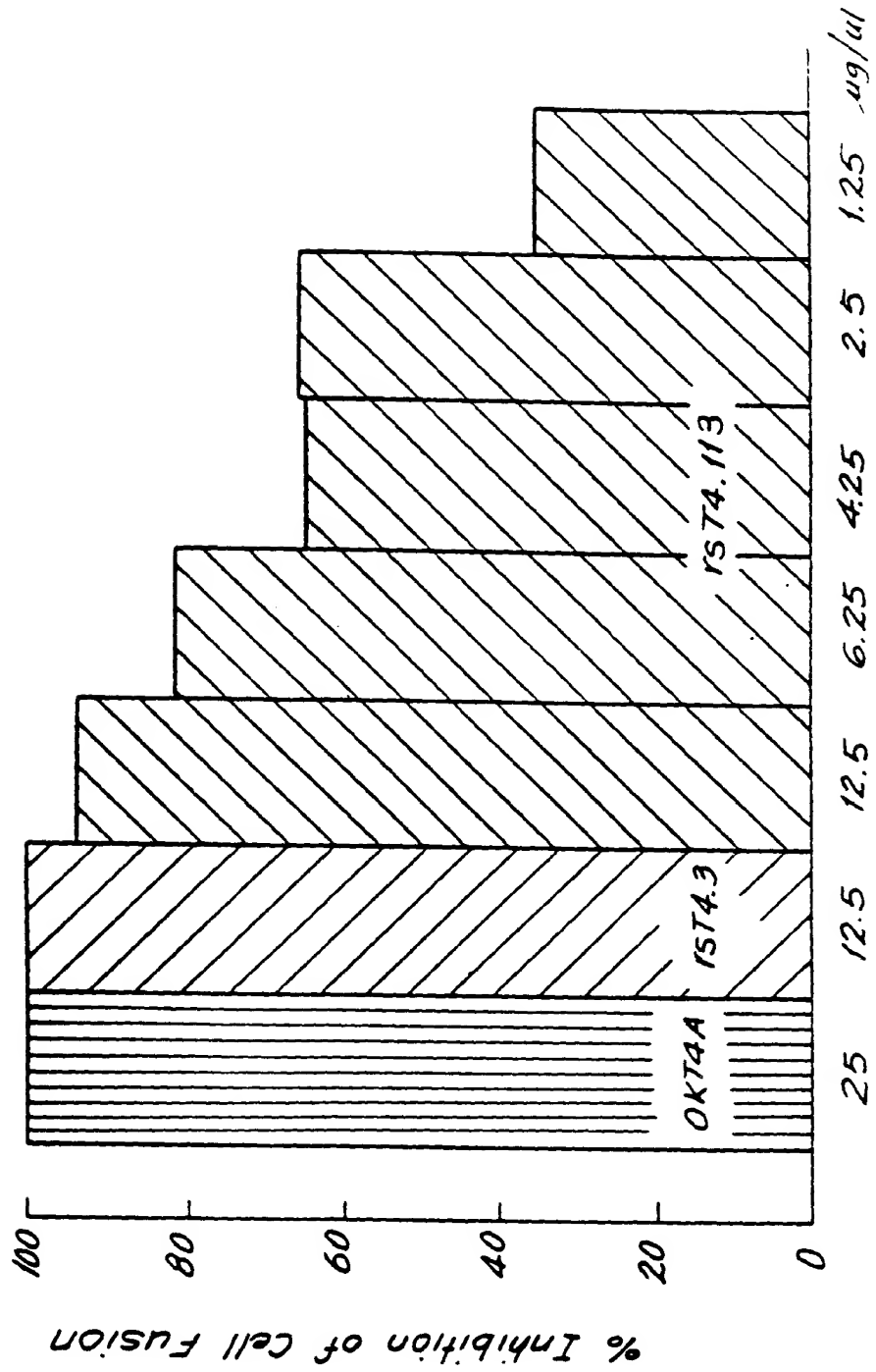
FIG. 42
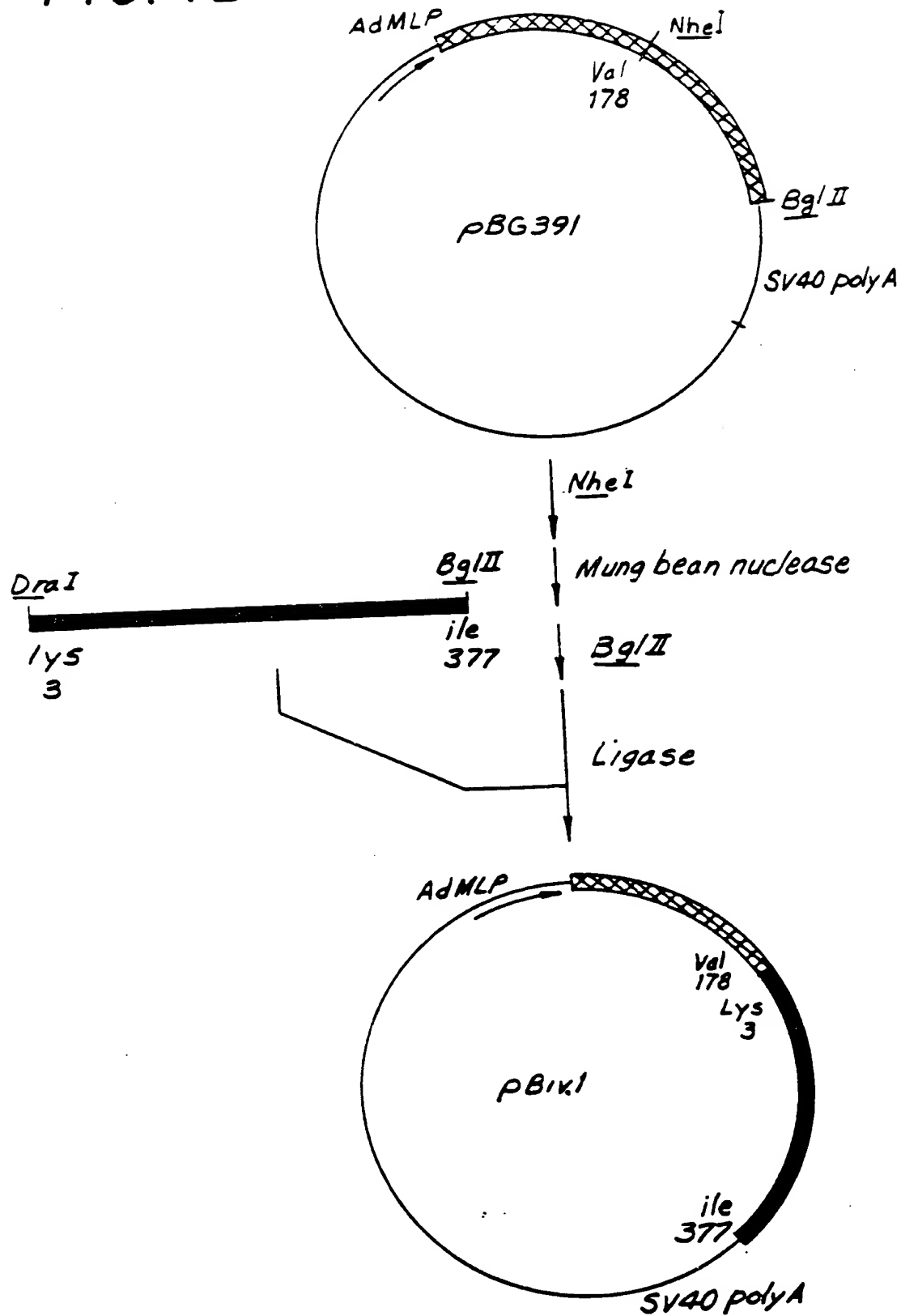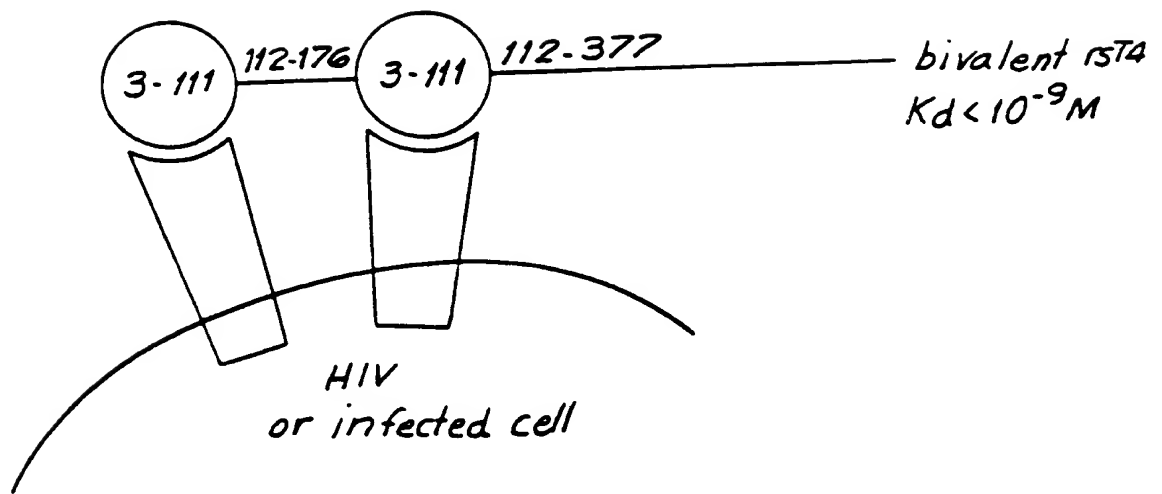
FIG.43

MICROORGANISMS

Original Sheet in connection with the microorganism referred to on page ___ 95, lines 29-35 ___ of the description.
96, lines 1-1

96, lines 19-21

A. IDENTIFICATION OF DEPOSIT

Further deposits are identified on an additional sheet ☑ — 3 additional sheets attached

Name of depositary institution

In Vitro International, Inc.

Address of depositary institution (including postal code and country)

611 (P) Hammonds
Ferry Road, Linthicum, Maryland 21090
United States of America

Date of deposit	Accession Number
See attached additional sheets	See attached additional sheets

B. ADDITIONAL INDICATIONS (leave blank if not applicable. This information is continued on a separate attached sheet ☐)

In respect of those designations in which a European patent is sought samples of the deposited microorganisms will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 28(4) EPC).

C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)

D. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later (Specify the general nature of the indications e.g. "Accession Number of Deposit")

E. ☐ This sheet was received with the international application when filed (to be checked by the receiving Office)

_____
(Authorized Officer)

☐ The date of receipt (from the applicant) by the International Bureau is

— 13 JANUARY 1989
(13. 01. 89)

J.-L. Baum _____
(Authorized Officer)

(January 1985)

Continuation Of Box A

IDENTIFICATION OF DEPOSITS

BG378: <u>E.coli</u> MC1061/pBG378
199-7: <u>E.coli</u> MC1061/p199-7
170-2: <u>E.coli</u> JA221/p170-2
EC100: <u>E.coli</u> JM83/pEC100
BG377: <u>E.coli</u> MC1061/pBG377
BG380: <u>E.coli</u> MC1061/pBG380
BG381: <u>E.coli</u> MC1061/pBG381

DATE OF DEPOSITS

2 September 1987

ACCESSION NUMBERS

IVI 10143
IVI 10144
IVI 10145
IVI 10146
IVI 10147
IVI 10148
IVI 10149

Continuation Of Box A

IDENTIFICATION OF DEPOSITS

BG-391:   E.coli MC1061/pBG391
BG-392:   E.coli MC1061/pBG392
BG-393:   E.coli MC1061/pBG393
BG-394:   E.coli MC1061/pBG394
BG-396:   E.coli MC1061/pBG396
203-5:   E.coli SG936/p203-5

DATE OF DEPOSITS

6 January 1988

ACCESSION NUMBERS

IVI 10151
IVI 10152
IVI 10153
IVI 10154
IVI 10155
IVI 10156

Continuation Of Box A

IDENTIFICATION OF DEPOSITS

211-11:   E.coli A89/pBG211-11
214-10:   E.coli A89/pBG214-10
215-7:   E.coli A89/pBG215-7

DATE OF DEPOSITS

24 August 1988

ACCESSION NUMBERS

IVI 10183
IVI 10184
IVI 10185

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US88/02941

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) ⁶

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(4): C07H 15/12; C12Q 1/70, C12Q 1/02, see attachment.
U.S. CL.: 536/27; 435/5,29,39,68,91,170,172.3, see attachment.

II. FIELDS SEARCHED

Minimum Documentation Searched ⁷

Classification System	Classification Symbols
U.S.	435/5,29,39,68,91,170,172.1,172.3,240.253,320; 530/350,412; 514/2; 424/85; 536/27; 935/6, 9, 11,12,15,22,23,24,59,60,65,66

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched ⁸

Chemical Abstract Data Base (CAS) 1967-1988; Biosis Data Base 1969-1988 Keywords: CD4, T4, TCell, AIDS, HTLV, HTLVI, HTLVIII, see attachment.

III. DOCUMENTS CONSIDERED TO BE RELEVANT ⁹

Category *	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
Y	SCIENCE, Volume 234, issued 1986 November, (Washington, DC., U.S.A.), (Q.J. SATTENTAU ET AL), "Epitopes of the CD4 Antigen and HIV Infection" See pages 1120-1123. See particularly page 1120	13-20, 29-33 and 48-52
Y	SCIENCE, Volume 234, issued 1986, November, (Washington, D.C. U.S.A) (J.A. HOXIE ET AL), "Alterations in T4 (CD4) Protein and mRNA synthesis in Cells Infected with HIV" see pages 1123-1127. See particularly page 1123.	13-20, 29-33, and 48-52
Y,P	PROCEEDINGS NATIONAL ACADEMY OF SCIENCES, U.S.A, Volume 84, issued 1987 December (Washington, D.C., U.S.A), (P.J. MADDON ET AL.), "Structure and Expression of the Human and Mouse T4 Genes", See pages 9155-9159, See particularly page 9155 and 9156.	1-4,25-27, 34-36 and 39-46

* Special categories of cited documents: ¹⁰

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
2? NOVEMBER 1988	0 3 FEB 1989
International Searching Authority	Signature of Authorized Officer
ISA/US	RICHARD C. PEET

Form PCT/ISA/210 (second sheet) (Rev.11-87)

III DOCUMENTS CONSIDERED TO BE RELEVANT   (CONTINUED FR M THE SECOND SHEET)

Category *	Citation of Document with indication where appropriate, of the relevant passages	Relevant to Claim No
Y	PROCEEDINGS NATIONAL ACADEMY OF SCIENCES, U.S.A, Volume 84, issued 1987, June (Washington, D.C., U.S.A.), (T.C. CHANH ET AL.), "Monoclonal Anti-idiotypic Antibody Mimics the CD4 Receptor and Binds Human Immunodeficiency Virus" See pages 3891-3895. See particularly page 3891.	13-24, 29-33 and 48-52
X̲ Y	CELL, Volume 47, issued 1986, November, (Cambridge, Mass., U.S.A) (P.J. MADDON ET AL), "The T4 Gene Encodes the AIDS Virus Receptor and is Expressed in the Immune System and the Brain", See pages 333-348, See particularly pages 333-335.	1,3-6 and 25-27 2,7-24 and 28-50
P,Y	CHEMICAL ABSTRACTS, Volume 107, no. 15, issued 1987 October 12 (Columbus, Ohio, U.S.A), T.L. LENTZ et al, "Rabies virus binding to cellular membranes measured by enzyme immunoassay' see page 359, column 1, the abstract no. 131853f, Muscle Nerve, 1985, 8(4), 336-345 (Eng).	16-18 32-33 and 50
Y	CHEMICAL ABSTRACTS, Volume 106, no. 21, issued 1987, May 25, (Columbus, Ohio, U.S.A), J.P. ZIMMER ET AL., 'Diphenylhydantoin (DPH) blocks HIV-receptor on T-lymphocyte surface', see page 123, column 1, the abstract no. 168522c, Blut, 1986, 53(6), 447-450 (Eng).	13-15, 19-20, 29-30, 48-49 and 51-52
Y,P	BIOLOGICAL ABSTRACTS, Volume 85, no. 4, issued 1988, April 15 (Philadelphia, PA, U.S.A), A.G. DALGLEISH ET AL., 'Neutralization of HIV isolates by anti-idiotypic antibodies which mimic the T4 (CD4) epitope: A potential AIDS vaccine' see page 222, abstract no. 37595, Lancet 2 (8567): 1047-1050 (Eng).	13-15, 19-20, 29-30, 48-49, and 51-52